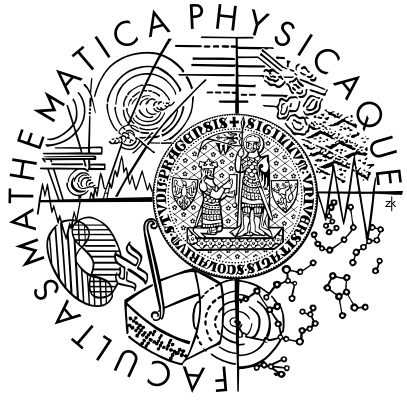
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Radka Picková

Odhady Value-at-Risk – nestandardní postupy

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: *Prof. RNDr. Jitka Dupačová, DrSc.*

Studijní program: *Matematika*

2008

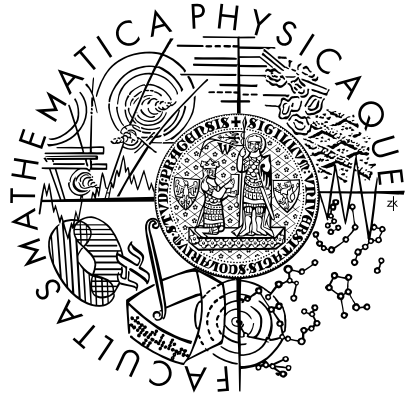
Poděkování:
Děkuji Prof. RNDr. Jitce Dupačové, DrSc. za její rady a cenné připomínky k obsahu práce. Děkuji Simone Manganelli za svolení pracovat s jeho CAViaR programy. Své rodině a přátelům děkuji za podporu.

Prohlašuji, že jsem svou diplomovou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 9. dubna 2008 Radka Picková

Charles University in Prague
Faculty of Mathematics and Physics

DIPLOMA THESIS



Radka Picková

Value-at-Risk estimation – non standard approaches

Department of Probability and Mathematical Statistics

Supervisor: *Prof. RNDr. Jitka Dupačová, DrSc.*

Study program: *Mathematics*

2008

Contents

1 Introduction to Value-at-Risk **6**
1.1 VaR – Summary 10

2 CAViaR **12**
2.1 CAViaR models 12
2.2 Regression quantiles 14
2.3 Evaluating VaR models 26
2.4 Empirical results 28

3 Filtered historical simulation **38**
3.1 Historical simulation 38
3.1.1 Empirical results 39
3.2 Filtered historical simulation 41
3.3 Volatility updating 44
3.3.1 Empirical results 46

4 Conclusion **49**

A Graphs and tables **50**

Bibliography **65**

Název práce: *Odhady Value-at-Risk – nestandardní postupy*
Autor: *Radka Picková*
Katedra: *Katedra pravděpodobnosti a matematické statistiky*
Vedoucí diplomové práce: *Prof. RNDr. Jitka Dupačová, DrSc.*
e-mail vedoucího: *dupacova@karlin.mff.cuni.cz*

Abstrakt:
Tématem předložené práce je Value-at-Risk (VaR) a její odhadování. VaR je finanční míra rizika definovaná jako kvantil rozdělení budoucích výnosů, resp. ztrát. Existuje mnoho metod s různými předpoklady, jak lze VaR odhadovat. Nejběžněji používané metody obvykle předpokládají, že výnosy, resp. ztráty, jsou nezávislé stejně rozdělené, především že jsou normálně rozdělené. Protože denní finanční data většinou tento předpoklad nesplňují, bylo navrženo mnoho alternativních postupů, jak odhadovat VaR. V předložené práci jsou detailněji popsány dvě metody, metoda CAViaR a její asymptotické vlastnosti a metoda filtrace historických simulací. Součástí práce jsou numerické experimenty s reálnými daty.

Klíčová slova: historické simulace, kvantilová regrese, Value-at-Risk.

Title: *Value-at-Risk estimation – non standard approaches*
Author: *Radka Picková*
Department: *Department of Probability and Mathematical Statistics*
Supervisor: *Prof. RNDr. Jitka Dupačová, DrSc.*
Supervisor's e-mail address: *dupacova@karlin.mff.cuni.cz*

Abstract:
The topic of the presented work is Value-at-Risk (VaR) and its estimation. VaR is a financial risk measure and is defined as a quantile of the distribution of future returns, resp. losses. There exist various methods based on different assumptions how to estimate VaR. The most commonly used methods usually assume that the returns, resp. losses, are independently and identically distributed, especially that they are normally distributed. Since this assumption is not satisfied for most daily financial data, many alternative approaches have been suggested to estimate VaR. In the presented work two of them are discussed in detail, the CAViaR method and its asymptotic properties and the method of filtered historical simulation. One part of the work are numerical experiments with real data.

Keywords: historical simulations, quantile regression, Value-at-Risk.

Chapter 1

Introduction to Value-at-Risk

Value-at-Risk (VaR) has become the standard measure that financial analysts use to quantify the downside market risk. VaR equals the loss on the portfolio of financial instruments that will not be exceeded by the end of the time period with a specified confidence level. In simpler words, it is a number that indicates how much a financial institution can lose with a given probability over a given time horizon. The great popularity that this instrument has achieved among financial practitioners is mainly caused by its simplicity: VaR reduces the market risk associated with any portfolio to just one number. VaR measures can have many applications, for example in risk management, to evaluate the performance of risk takers and for regulatory requirements, and therefore it is very important to be able to estimate VaR accurately enough.

VaR models measure the market risk by determining how much the value of the portfolio could decline as a result of changes in market prices or rates. The two most important components of VaR models are the length of the time horizon over which the market risk is measured and the confidence level at which the market risk is measured. The choice of these components by risk managers affects the choice of suitable VaR models.

The time horizon used in the definition of VaR, often referred to as the holding period, can be chosen arbitrarily. In practice, it varies from one day to two weeks (ten trading days) and depends on liquidity of assets and frequency of trading transactions. VaR models usually assume that the portfolio's composition does not change over the holding period. This assumption argues for the use of short holding periods because the composition of active trading portfolios is usually changing frequently. Therefore, most of the articles concerning VaR focuses on the widely used one-day holding period. The Basel Committee on Banking Supervision recommends to use the ten-day holding period. Some practitioners dispute that the time horizon of ten days is inadequate for frequently traded instruments and is restrictive for illiquid assets. Longer holding periods are usually recommended for portfolios with illiquid instruments.

VaR measures are often expressed as percentiles corresponding to the desired confidence level. Here, we must not confuse the confidence level and the given

probability of possible exceeding of the VaR threshold. For example, an estimate of risk at the 99 percent confidence level is the amount of loss that a portfolio is expected to exceed only 1 percent of the time. We will call such an amount the 1% VaR. However, in some articles it is also possible to find the term 99th percentile VaR measure because the amount is the 99th percentile of the distribution of potential losses on the portfolio. One must decide whether to work with the distribution of losses or with the distribution of returns which designates whether one estimates the right or the left tail of the distribution, see Figure 1.1 for better understanding. Note that losses are equal to negative returns. VaR is a number such that losses will exceed it with a small probability $\delta\,\%$ (e.g. $5\,\%$, $1\,\%$ etc.), therefore VaR can be computed as the $(100-\delta)\%$ quantile of the distribution of losses or as the negative $\delta\%$ quantile of the distribution of returns, as illustrated in Figure 1.1. From now on we will always consider the distribution of returns. In practice, VaR estimates are calculated from the 10 to 0.01 percent, but the most commonly used range is the 5 to 1 percent range. The choice of the probability level can have a substantial effect on the performance of VaR approaches. Note that regulators have chosen the probability level equal to $1\,\%$.

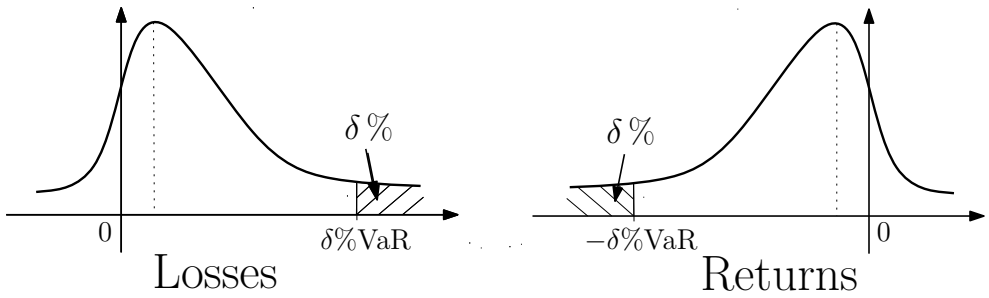


Figure 1.1: VaR – Density of probability distribution of losses on the left, resp. returns on the right.

Regarding confidence levels and probabilities here is an example how the term VaR is not used correctly. A wrong definition that is often used for the one day 5% VaR is the following: VaR is the *maximum* potential loss that a portfolio can suffer in the $5\,\%$ worst cases in one day. Corrected version (which may sound odd): VaR is the *minimum* potential loss that a portfolio can suffer in the $5\,\%$ worst cases in one day. Politically correct version: VaR is the *maximum* potential loss that a portfolio can suffer in the $95\,\%$ worst cases in one day.

Now we will define VaR precisely. As we mentioned above, VaR is a number such that losses will exceed the VaR threshold with only a small probability θ, typically chosen between $1\,\%$ and $5\,\%$. More specifically, conditional on the information given up to time t, the θ VaR at time $t+h$ of one unit of investment is the negative θ-quantile of the conditional return distribution, that is

$$\mathrm{VaR}^{\theta}_{t+h} := -Quant_{\theta}(y_{t+h} \mid \Omega_t) = -\inf_{x}\big\{x \in \mathbb{R} : \mathsf{P}\big(y_{t+h} \le x \mid \Omega_t\big) \ge \theta\big\}, 0 < \theta < 1,$$

where $Quant_{\theta}(\cdot)$ denotes the quantile function, y_t is the return on an asset or

portfolio in period t, and Ω_t represents the information available at time t. We subsequently suppress superscript θ for simplicity.

Even though VaR has frequently been criticized as a risk measure in last few years (for more details see section 1.1), regulators still use VaR as the standard risk measure to assess capital requirements. Due to this practical relevance of the VaR concept, the need for reliable VaR estimation and prediction strategies arises. Despite the fact that VaR is an easy and intuitive concept, its measurement is a challenging statistical problem. From a statistical point of view, VaR measure involves the estimation of a quantile of the distribution of future returns.

There are various methods how to deal with VaR estimation. If we actually know the precise distribution of future returns and the distribution function F is increasing we can simply compute VaR as the negative θ-quantile of the known return distribution, that is θ VaR is $-F^{-1}(\theta)$. Moreover, if we assume that the distribution of random returns Y belongs to the location-scale family, which means that the distribution function is

$$F_{\mu,\sigma}(y) = G\left(\frac{y-\mu}{\sigma}\right),$$

where μ is the parameter of location, $\sigma > 0$ is the scale parameter and $G(y)$ is a distribution function with zero mean and unit variance, then the following holds

$$\mathsf{P}(Y \leq -\mathrm{VaR}_\theta) = \mathsf{P}\left(\frac{Y-\mu}{\sigma} \leq \frac{-\mathrm{VaR}_\theta - \mu}{\sigma}\right) = G\left(\frac{-\mathrm{VaR}_\theta - \mu}{\sigma}\right) = \theta$$

and we get that

$$\mathrm{VaR}_\theta = -\mu - \sigma q_\theta(Y) = -\big(\mu + \sigma q_\theta(Y)\big),$$

where q_θ is the θ-quantile of the distribution function G. Note that in the special case when returns are normally distributed with zero expected value VaR is a multiple of the standard deviation σ. That may have been one reason why the concept of VaR was naturally accepted since the standard deviation itself was (and also is) sometimes used as a measure of risk.

If we do not know the precise future distribution of returns we need to estimate VaR using historical data (returns in the past). One possibility is the parametric approach, that means to estimate the type of the distribution on the basis of the empirical distribution function. Once we choose the type of distribution of returns, we estimate the unknown parameters, such as the mean, the standard deviation (in the case of normal distribution) and some additional shape parameters (for instance degrees of freedom in the case of Student's t-distribution). Then the estimator of VaR is the negative θ-quantile of the estimated distribution of returns.

Another possibility is the nonparametric approach which makes no distributional assumptions about portfolio returns. The nonparametric historical simulation

simply utilizes empirical quantiles based on the available past data, which are assumed to be independent and identically distributed (i.i.d.). Consider that we have n observations of past returns, we sort them in ascending order, then the empirical θ-quantile is given by the return that leaves $\theta\,\%$ of the observations on its left side and $(1-\theta)\,\%$ on its right side. If such a number falls between two consecutive returns, then some interpolation rule is applied. The historical simulation is one of the simplest methods and can be applied to complicated financial instruments since we do not need to know the distribution. However, the assumption of i.i.d.-ness of returns is often violated and the accuracy of the estimator depends on the choice of θ and the amount of available past data (smaller θ needs enough data).

In practise, VaR estimation is not so simple as it may seem. Empirical experience shows that financial return distributions are not constant over time and exhibit nonstandard statistical properties. Specifically, they are not independently and identically distributed and, moreover, they are not normally distributed in most cases. This is reflected by three widely reported empirical facts about financial markets[1]:

- volatility clustering, indicated by significant autocorrelation of absolute and squared returns,
- substantial kurtosis, that is, the density of the unconditional return distribution is more peaked around the center and possesses much fatter tails than the normal density,
- skewness of returns (equity returns are typically negatively skewed).

As a consequence, standard methods, based on the assumption of i.i.d.-ness and normality, are not adequate in many situations, which has led to various alternative strategies for VaR prediction.

Note that the high positive kurtosis means that extreme returns are more likely than the normal distribution would predict. When VaR is estimated with the assumption of normally distributed returns and the returns exhibit fatter tails than the normal distribution, then the actual VaR is underestimated. Therefore, numerous different distributions (with fat tails and possibly skewed) have been suggested as substitutes for the normal distribution, for instance the stable Paretian distribution[2] (see Rachev and Mittnik (2000)), the Student's t-distribution (see e.g. Huisman et al. (1998)), the skewed t-distribution (see Kuester et al. (2006) and the references therein) or a mixture of normal distributions (see Haas et al. (2004)).

One alternative approach is the extreme value theory. Unlike the fully parametric models which describe the entire distribution of returns (including possible volatility dynamics), the semiparametric extreme value theory focuses on modeling the

[1] For instance see Duffie and Pan (1997).

[2] Note that the normal distribution is a special case of stable Paretian distributions.

tails of the return distribution. This method has showed very good results when modeling extreme quantiles, such as the 1, 0.1 or 0.01 percent VaR. For more details regarding extreme value theory see for example Danielsson and de Vries (2000), where a method using tail estimators is discussed.

Other alternative approaches are the quantile regression which models a specific quantile rather than the whole distribution of returns and the filtered historical simulation which is an extension of the nonparametric historical simulation. We will describe these methods in the following chapters in more details.

Various models for calculating VaR are reviewed in Duffie and Pan (1997) and numerous alternative strategies are compared in Kuester et al. (2006).

Coherent Risk Measures

Artzner et al. (1999) were the first who stated convenient properties which should any risk measure have and they defined coherent measures of risk.

Consider a real-valued random variable X on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ to be the random return (or loss for negative values) of some asset or portfolio.

Consider a set V of real-valued random variables (representing random returns) on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that $\mathsf{E}[X^-] < \infty$ for all $X \in V$. The function $\rho : V \to \mathbb{R}$ is a coherent risk measure if it satisfies the following four axioms:

1. monotonicity: for all X and $Y \in V$ with $X(\omega) \leq Y(\omega),\ \forall\, \omega \in \Omega$, we have $\rho(X) \geq \rho(Y)$,
2. subadditivity: $X,\ Y,\ X + Y \in V\ \Rightarrow\ \rho(X + Y) \leq \rho(X) + \rho(Y)$,
3. positive homogeneity: $X \in V,\ h > 0,\ hX \in V\ \Rightarrow\ \rho(hX) = h\rho(X)$,
4. translation invariance: $X \in V,\ a \in R\ \Rightarrow\ \rho(X + a) = \rho(X) - a$.

The axiom of monotonicity can be rewritten as

1b. $X \in V,\ X(\omega) \geq 0,\ \forall\, \omega \in \Omega\ \Rightarrow\ \rho(X) \leq 0$.

Subadditivity together with homogeneity implies

5. convexity: $X,\ Y \in V,\ \alpha \in [0, 1]\ \Rightarrow\ \rho\big(\alpha X + (1-\alpha)Y\big) \leq \alpha\rho(X) + (1-\alpha)\rho(Y)$.

1.1 VaR – Summary

As we have mentioned above, VaR methodologies have become customary tools in risk management. Therefore, it is important to be aware of VaR strengths and weaknesses which we will now briefly summarize.

Advantages and Usage:

- VaR is a general-purpose measure of risk (it can be applied to any kind of risk and aggregates different risks into one number),
- VaR usually includes an estimate of future events,
- VaR is widely used by financial institutions, fund managers, and nonfinancial corporations to control the market risk in a portfolio of financial instruments,
- VaR is applied for allocating financial resources and risk-adjusted performance evaluation,
- VaR has been adopted by central bank regulators as the major determinant of the capital banks are required to keep to cover potential losses arising from the market risks they are bearing,
- banks may use their VaR models for calculation of regulatory market-risk capital requirements.

Disadvantages:

- VaR is not a coherent risk measure (VaR fails to satisfy the subadditivity property except for the elliptic distributions[3]), see Artzner et al. (1999),
- VaR does not say how severe the loss exceeding VaR can be,
- VaR does not differentiate between shapes of distributions,
- VaR is not convex, therefore it is not easy to optimize the portfolio composition with respect to minimal VaR[4] (problems with local extremes),
- VaR does not behave nicely with respect to addition of risks, even independent ones, creating severe aggregation problems,
- VaR can fail to appropriately account for portfolio risk diversification.

The remaining text is organized as follows. In chapter 2, the CAViaR method, nonlinear regression quantile techniques and methods for evaluating VaR models are described and followed by empirical results. In chapter 3, methods of historical simulation and filtered historical simulation are described and compared on the basis of empirical results. Chapter 4 contains concluding remarks. Additional figures and tables are given in the appendix.

[3]Elliptic distributions (of which the normal is a special case) are distributions whose density is constant on ellipsoids. The density function of a multivariate elliptic distribution with parameters $(\boldsymbol{\mu}, \Sigma)$ has the form $f(\mathbf{x}) = |\Sigma|^{-1/2}\, g\big((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\big)$, where $\boldsymbol{\mu} \in \mathbb{R}^p$, Σ is a $p \times p$ positive definite matrix, $g(u), u \geq 0$ is a univariate function such that $\int_0^\infty u^{p/2-1} g(u) du = \frac{\Gamma[p/2]}{\pi^{p/2}}$ and Γ is the gamma function (see He and Wang (1997)).

[4]See Gaivoronski and Pflug (2005) where applications of the VaR concept in the context of optimal portfolio selection are discussed.

Chapter 2

CAViaR

Engle and Manganelli (2004) propose a new approach to quantile estimation. They suggest to model directly the evolution of the quantile over time, instead of modeling the whole distribution of portfolio returns. The reason why they propose this new approach is the empirical fact that volatilities of stock market returns cluster over time. This fact may be translated in statistical words by saying that the distribution of stock market returns is autocorrelated. Consequently, the VaR, which is tightly linked to the standard deviation of the distribution, must exhibit similar behavior. A natural way to formalize this characteristic is to use some type of autoregressive specification. Therefore, they propose a conditional autoregressive quantile specification, which they call Conditional Autoregressive Value at Risk (CAViaR).

2.1 CAViaR models

Suppose that we observe a vector of portfolio returns, $\{y_t\}_{t=1}^{T}$. Let θ be the probability associated with VaR, let $\mathbf{x}_t$ be a vector of time t observable variables, and let $\boldsymbol{\beta}_\theta$ be a p-vector of unknown parameters. Let $f_t(\boldsymbol{\beta}) \equiv f_t(\mathbf{x}_{t-1}, \boldsymbol{\beta}_\theta)$ denote the time t θ-quantile of the distribution of portfolio returns formed at time $t-1$ and let $\mathrm{VaR}_t(\boldsymbol{\beta}) \equiv \mathrm{VaR}_t(\mathbf{x}_{t-1}, \boldsymbol{\beta}_\theta)$ denote the negative θ-quantile of the distribution of portfolio returns at time t formed at time $t-1$, that is the one day θ VaR prediction for time t (the θ subscript is suppressed for notational convenience). Note that $\mathrm{VaR}_t(\boldsymbol{\beta}) = -f_t(\boldsymbol{\beta})$. The basic idea is to model the quantile as some function of the past information. Engle and Manganelli (2004) propose the following generic CAViaR specification

$$\mathrm{VaR}_t(\boldsymbol{\beta}) = \beta_0 + \sum_{i=1}^{q} \beta_i \mathrm{VaR}_{t-i}(\boldsymbol{\beta}) + \sum_{j=1}^{r} \beta_{q+j} l(\mathbf{x}_{t-j}), \tag{2.1}$$

where $p = q + r + 1$ is the dimension of $\boldsymbol{\beta}$ and l is a function of finite number of lagged values of observable variables. The autoregressive terms $\beta_i \mathrm{VaR}_{t-i}(\boldsymbol{\beta})$, $i = 1, \ldots, q$, ensure that the quantile changes smoothly over time. The role of $l(\mathbf{x}_{t-j})$ is to link $\mathrm{VaR}_t(\boldsymbol{\beta})$ to observable variables that belong to the information set. Various models can be estimated by choosing different specifications for the l function and by choosing particular observable variables for $\mathbf{x}_t$. A natural choice

for $\mathbf{x}_t$ is lagged returns, since we would like to link the conditional quantile to return innovations.

The choice of the best functional form is mainly an empirical problem and should be determined by the data set under study. Engle and Manganelli (2004) propose some specific examples of CAViaR processes. They suggest that one would expect VaR to increase as y_{t-1} becomes very negative, because one bad day makes the probability of the next somewhat greater. Moreover, they suggest that very good days might also increase VaR, as would be the case for volatility models. Hence, VaR could depend symmetrically on $|y_{t-1}|$. We use the notation $(x)^+ = \max(x, 0)$, $(x)^- = -\min(x, 0)$.

The first CAViaR specification introduced by Engle and Manganelli (2004) is

- Adaptive:

$$\mathrm{VaR}_t(\beta_1) = \mathrm{VaR}_{t-1}(\beta_1) + \beta_1 \left\{ \left[1 + \exp\big(\kappa[y_{t-1} + \mathrm{VaR}_{t-1}(\beta_1)]\big) \right]^{-1} - \theta \right\}, \quad (2.2)$$

where κ is some positive finite number. Note that as $\kappa \to \infty$, the last term converges almost surely to $\beta_1 \big\{ I\big(y_{t-1} \leq -\mathrm{VaR}_{t-1}(\beta_1)\big) - \theta \big\}$, and the adaptive specification can then be rewritten as

$$\mathrm{VaR}_t(\beta_1) = \mathrm{VaR}_{t-1}(\beta_1) + \beta_1 \Big[I\big(y_{t-1} \leq -\mathrm{VaR}_{t-1}(\beta_1)\big) - \theta \Big],$$

where $I(\cdot)$ represents the indicator function. For finite κ, this model is a smoothed version of a step function. The adaptive model incorporates the following simple rule: whenever the VaR is exceeded one should immediately increase it, but when the VaR is not exceeded, one should decrease it very slightly. This strategy will obviously reduce the probability of sequences of VaR exceeding (violation clusters) and will also make it unlikely that the VaR will never be exceeded. However, it does not make a difference between returns that are close to the VaR threshold and returns that extremely positive, when κ is large. It increases the VaR by the same amount regardless of whether the losses exceeded the VaR by a small margin or a large margin. The adaptive model (2.2) has a unit coefficient on the lagged VaR. The next CAViaR specification is:

- Symmetric absolute value:

$$\mathrm{VaR}_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 \, \mathrm{VaR}_{t-1}(\boldsymbol{\beta}) + \beta_3 \, |y_{t-1}|, \quad (2.3)$$

which allows the autoregressive parameter β_2 to be different from one. This specification introduces a direct response of the quantile to the return process and responds symmetrically to past extreme returns. The assumption of symmetry is relaxed in the next CAViaR specification:

- Asymmetric slope:

$$\mathrm{VaR}_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 \, \mathrm{VaR}_{t-1}(\boldsymbol{\beta}) + \beta_3 \, (y_{t-1})^+ + \beta_4 \, (y_{t-1})^-. \quad (2.4)$$

The asymmetric slope specification (2.4) differs from the symmetric absolute value specification (2.3) in that it allows the response to positive and negative returns to be different. Another alternative is:

- Indirect GARCH(1,1):

$$\mathrm{VaR}_t(\boldsymbol{\beta}) = \left(\beta_1 + \beta_2 \,\mathrm{VaR}_{t-1}^2(\boldsymbol{\beta}) + \beta_3 \, y_{t-1}^2\right)^{1/2}, \tag{2.5}$$

which like the symmetric absolute value (2.3) responds symmetrically to past returns. Unlike the adaptive specification (2.2) the symmetric absolute value (2.3), the asymmetric slope (2.4) and the indirect GARCH(1,1) (2.5) specifications do not constrain the coefficient on the lagged VaR to be one.

The indirect GARCH model (2.5) would be correctly specified if the underlying data were truly a GARCH(1,1) with an i.i.d. error distribution (see Bollerslev (1986)). Then it is useful for Monte Carlo simulation which allows to examine the properties of the CAViaR method. If the model (2.5) is correctly specified, then instead of using the CAViaR method it may be more efficient to estimate the GARCH model directly by maximum likelihood and then infer the VaR from the distribution of the standardized residuals. The symmetric absolute value (2.3) and asymmetric slope (2.4) quantile specification would be correctly specified by a GARCH process in which the standard deviation, rather than the variance, is modeled either symmetrically or asymmetrically with i.i.d. errors.

Kuester et al. (2006) propose another possible CAViaR specification. They suggest that autocorrelation in financial returns is often nonnegligible and that this property can be incorporated by extending the existing CAViaR framework by allowing the returns to have a time-varying mean which may be captured by a regression, ARMA, or other models. They choose an AR(1) model for the mean, $y_t = ay_{t-1}+\epsilon_t$, and specify a new model, the indirect AR(1)-GARCH(1,1) CAViaR

$$\mathrm{VaR}_t(\boldsymbol{\beta}) = -ay_{t-1} + \Big(\beta_0 + \beta_1\big(\mathrm{VaR}_{t-1}(\boldsymbol{\beta}) + ay_{t-2}\big)^2 + \beta_2\big(y_{t-1} - ay_{t-2}\big)^2\Big)^{1/2}.$$

In their application to real data they demonstrate that this more general CAViaR specification leads to a significant improvement in performance. For more details see Kuester et al. (2006).

The CAViaR specifications are more general than the aforementioned GARCH models. Various forms of non-i.i.d. error distributions can be modeled in this way. Actually, these models can be used for situations with constant volatilities but changing error distributions, or situations in which both error densities and volatilities are changing.

2.2 Regression quantiles

We need to choose a method to estimate the unknown parameters of CAViaR models. Engle and Manganelli (2004) suggest to estimate the parameters of CAViaR

models using nonlinear regression quantile techniques, utilizing regression quantiles which where introduced by Koenker and Bassett (1978).

Koenker and Bassett (1978) showed how to extend the notion of a sample quantile to a linear regression model. Let $y_1, \ldots, y_T$ be a sample of observations. As shown, for example, in Koenker (2005), the unconditional sample θ-quantile, $\theta \in (0,1)$, can be found as a solution to

$$\min_{\beta \in \mathbb{R}} \left\{ \sum_{t:\ y_t \geq \beta} \theta \mid y_t - \beta \mid + \sum_{t:\ y_t < \beta} (1-\theta) \mid y_t - \beta \mid \right\}.$$

Koenker and Bassett (1978) extended this to the classical linear regression framework. Consider a sample of observations $y_1, \ldots, y_T$ generated by the model

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}^0 + \varepsilon_{t\theta}, \quad Quant_\theta(\varepsilon_{t\theta} | \mathbf{x}_t) = 0,$$

where $\mathbf{x}_t$ is a p-vector of regressors (nonrandom vectors) and $Quant_\theta(\varepsilon_{t\theta}|\mathbf{x}_t)$ is the θ-quantile of the error term $\varepsilon_{t\theta}$ conditional on $\mathbf{x}_t$. Koenker and Bassett (1978) define the θth regression quantile estimator by

$$\hat{\boldsymbol{\beta}}(\theta) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \sum_{t:\ y_t \geq \mathbf{x}_t'\boldsymbol{\beta}} \theta \mid y_t - \mathbf{x}_t'\boldsymbol{\beta} \mid + \sum_{t:\ y_t < \mathbf{x}_t'\boldsymbol{\beta}} (1-\theta) \mid y_t - \mathbf{x}_t'\boldsymbol{\beta} \mid \right\}.$$

The key assumption in the linear quantile regression model is that $y_t = \mathbf{x}_t'\boldsymbol{\beta}^0 + \varepsilon_{t\theta}$. Note that the distribution of the error term is left unspecified. Consequently, the only assumption made is that the conditional quantile function is given by $Quant_\theta(y_t \mid \mathbf{x}_t) = \mathbf{x}_t'\boldsymbol{\beta}^0$, and thus $Quant_\theta(\varepsilon_{t\theta} \mid \mathbf{x}_t) = 0$.

A special case of regression quantiles is the least absolute deviation (LAD) model. It is well known that whenever the errors have a fat-tailed distribution then the LAD estimator is more robust than ordinary least squares (OLS) estimators. Koenker and Bassett (1978) ran a simple Monte Carlo experiment and showed how the empirical variance of the median, compared with the variance of the mean, is slightly higher under the normal distribution but much lower under all other distributions taken into consideration (Gaussian mixture, Laplace and Cauchy distributions).

Let $f_t(\boldsymbol{\beta})$ be a function of parameters $\boldsymbol{\beta}$. The θth regression quantile estimator is defined as any $\hat{\boldsymbol{\beta}}$ that solves

$$\min_{\beta} \frac{1}{T} \left\{ \sum_{t:\ y_t \geq f_t(\boldsymbol{\beta})} \theta \mid y_t - f_t(\boldsymbol{\beta}) \mid + \sum_{t:\ y_t < f_t(\boldsymbol{\beta})} (1-\theta) \mid y_t - f_t(\boldsymbol{\beta}) \mid \right\}. \tag{2.6}$$

If the function f_t is linear, that is $f_t(\boldsymbol{\beta}) \equiv \mathbf{x}_t'\boldsymbol{\beta}$, then we can apply the results of Koenker and Bassett (1978). The question is what happens if the function f_t is not linear in the parameters $\boldsymbol{\beta}$. Then there are different possibilities how to solve the minimization problem with nonlinear objective function (2.6). Firstly, there exist

algorithms to minimize nonlinear objective function but these are not easy to implement. Secondly, we can rewrite the problem in the form of minimization with linear objective function and nonlinear conditions. However, the absolute value causes that we minimize a non-differentiable objective function. We can rewrite the absolute value in the objective function using the notation $|x| = (x)^+ + (x)^-$.

Note that

$$\begin{aligned}
&\min_{\beta} \frac{1}{T} \left\{ \sum_{t:\ y_t \geq f_t(\boldsymbol{\beta})} \theta \ \big|y_t - f_t(\boldsymbol{\beta})\big| + \sum_{t:\ y_t < f_t(\boldsymbol{\beta})} (1-\theta) \ \big|y_t - f_t(\boldsymbol{\beta})\big| \right\} \\
&= \min_{\beta} \frac{1}{T} \Bigg\{ \sum_{t=1}^{T} \Big[I\big(y_t \geq f_t(\boldsymbol{\beta})\big) \ \theta \ \big|y_t - f_t(\boldsymbol{\beta})\big| \\
&\qquad\qquad + \ I\big(y_t < f_t(\boldsymbol{\beta})\big) \ (1-\theta) \ \big|y_t - f_t(\boldsymbol{\beta})\big| \ \Big] \Bigg\} \\
&= \min_{\beta} \frac{1}{T} \sum_{t=1}^{T} \Big\{ \Big[I\big(y_t \geq f_t(\boldsymbol{\beta})\big) \ \theta \ - \ I\big(y_t < f_t(\boldsymbol{\beta})\big) \ (1-\theta) \Big] \big[y_t - f_t(\boldsymbol{\beta})\big] \Big\} \\
&= \min_{\beta} \frac{1}{T} \sum_{t=1}^{T} \Big\{ \Big[I\big(y_t \geq f_t(\boldsymbol{\beta})\big) \ \theta \ + \ I\big(y_t < f_t(\boldsymbol{\beta})\big) \ (\theta-1) \Big] \big[y_t - f_t(\boldsymbol{\beta})\big] \Big\} \\
&= \min_{\beta} \frac{1}{T} \sum_{t=1}^{T} \Big\{ \Big[\theta \ - \ I\big(y_t < f_t(\boldsymbol{\beta})\big) \Big] \big[y_t - f_t(\boldsymbol{\beta})\big] \Big\}
\end{aligned}$$

Therefore, (2.6) can be rewritten as

$$\min_{\beta} \frac{1}{T} \sum_{t=1}^{T} \Big\{ \Big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big) \Big] \big[y_t - f_t(\boldsymbol{\beta})\big] \Big\}. \tag{2.7}$$

The great advantage of this approach is that the only assumption required under this framework is that the quantile process is correctly specified. In particular, no distributional assumptions for the time series behavior of returns are needed, respectively no assumption on the distribution of the error terms is needed, hence reducing the risk of misspecification.

Consider the model

$$\begin{aligned}
y_t &= f(y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1; \boldsymbol{\beta}^0) + \varepsilon_{t\theta} \quad \big[Quant_\theta(\varepsilon_{t\theta}|\Omega_t) = 0\big] \\
&\equiv f_t(\boldsymbol{\beta}^0) + \varepsilon_{t\theta}, \qquad t = 1, \ldots, T,
\end{aligned} \tag{2.8}$$

where $f_1(\boldsymbol{\beta}^0)$ is some given initial condition, $\mathbf{x}_t$ is a vector of exogenous or predetermined variables, $\boldsymbol{\beta}^0 \in B$ is the vector of true unknown parameters that need to be estimated, B is a compact subset of $\mathbb{R}^p$, and $\Omega_t = \big[y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1; f_1(\boldsymbol{\beta}^0)\big]$ is the information set available at time t. Let $\hat{\boldsymbol{\beta}}_T$ be the parameter vector that minimizes the objective function in equation (2.7) over the set B.

Theorems 1 and 2 below show that the nonlinear regression quantile estimator $\hat{\boldsymbol{\beta}}_T$ is consistent and asymptotically normal if the function f_t in (2.8) satisfies stated conditions and technical assumptions. The Theorem 3 below provides a consistent estimator of the variance-covariance matrix.

In the following we denote the conditional density of $\varepsilon_{t\theta}$ evaluated at 0 by $h(0|\Omega_t)$, the $1 \times p$ gradient of $f_t(\boldsymbol{\beta})$ by $\nabla f_t(\boldsymbol{\beta})$, and define $\nabla f(\boldsymbol{\beta})$ to be a $T \times p$ matrix with typical row $\nabla f_t(\boldsymbol{\beta})$.

Theorem 1 (Consistency). Consider the model (2.8). Let $\hat{\boldsymbol{\beta}}_T$ be a solution to

$$\min_{\boldsymbol{\beta} \in B} T^{-1} \sum_{t=1}^{T} \Big\{ \Big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big)\Big] \big[y_t - f_t(\boldsymbol{\beta})\big] \Big\}. \tag{2.9}$$

Let the following assumptions be satisfied:

C0. $(\Omega, \mathcal{F}, \mathsf{P})$ is a complete probability space and $\{\varepsilon_{t\theta}, \mathbf{x}_t\}, t = 1, 2, \ldots,$ are random vectors on this space.

C1. The function $f_t(\boldsymbol{\beta}) : \mathbb{R}^{k_t} \times B \to \mathbb{R}$ is such that for each $\boldsymbol{\beta} \in B$, a compact subset of $\mathbb{R}^p$, $f_t(\boldsymbol{\beta})$ is measurable with respect to the information set Ω_t and $f_t(\cdot)$ is continuous in $B, t = 1, 2, \ldots,$ for a given choice of explanatory variables $\{y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1\}$.

C2. Conditional on all of the past information Ω_t, the error terms $\varepsilon_{t\theta}$ form a stationary process, with continuous conditional density $h_t(\varepsilon|\Omega_t)$.

C3. There exists $h > 0$ such that for all t, $h_t(0|\Omega_t) \geq h$.

C4. $|f_t(\boldsymbol{\beta})| < K(\Omega_t)$ for each $\boldsymbol{\beta} \in B$ and for all t, where $K(\Omega_t)$ is some (possibly) stochastic function of variables that belong to the information set Ω_t, such that $\mathsf{E}(|K(\Omega_t)|) \leq K_0 < \infty$, for some constant K_0.

C5. $\mathsf{E}[|\varepsilon_{t\theta}|] < \infty$ for all t.

C6. $\big\{ \big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big)\big] \big[y_t - f_t(\boldsymbol{\beta})\big] \big\}$ obeys the uniform law of large numbers[1].

C7. For every $\xi > 0$, there exists a $\tau > 0$ such that if $\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \geq \xi$, then $\liminf_{T\to\infty} T^{-1} \sum \mathsf{P}\big[|f_t(\boldsymbol{\beta}) - f_t(\boldsymbol{\beta}^0)| > \tau\big] > 0$.

Then

$$\hat{\boldsymbol{\beta}}_T \xrightarrow{\mathsf{P}} \boldsymbol{\beta}^0 \quad \text{as} \quad T \to \infty,$$

where the symbol $\xrightarrow{\mathsf{P}}$ denotes the convergence in probability.

Note that the Assumption C1, requiring continuity in the vector of parameters $\boldsymbol{\beta}$ of the quantile specification, is clearly satisfied by all of the CAViaR models considered in section 2.1 where we apply $f_t(\boldsymbol{\beta}) = -\mathrm{VaR}_t(\boldsymbol{\beta})$. Assumptions C3 and

[1] We refer to the Appendix 2 in White (1994) where the uniform law of large numbers is stated.

C7 are identification conditions that are common in the regression quantile literature. Assumptions C4 and C5 are dominance conditions that rule out explosive behavior.

In the proof of the Theorem 1, we apply the Corollary 5.12 of White (1994), which establishes an analog of the consistency result in Koenker and Bassett (1978) for nonlinear models of regression quantiles in a dynamic context. We will state the Corollary 5.12 of White (1994) before we start the proof of Theorem 1.

Assume that $(\Omega, \mathcal{F}, \mathsf{P})$ is a complete probability space and the observed data are a realization of a stochastic process $X \equiv \{X_t : \Omega \to \mathbb{R}^\nu, \nu \in \mathbb{N}, t = 1, 2, \ldots\}$ on this space, which means that $X_t, t = 1, 2, \ldots,$ are random vectors on this space. We usually do not have available observations on an entire infinite sequence of realizations. Instead we have a realization $\mathbf{x}^T$ of a finite history, $X^T \equiv (X_1', \ldots, X_T')'$. We call $\mathbf{x}^T$ a sample of size T.

Assume we have a collection of sequences of functions

$$g(\boldsymbol{\beta}) \equiv \left\{ g_t(\cdot, \boldsymbol{\beta}) : \mathbb{R}^{\nu t} \to \mathbb{R}^+, t = 1, 2, \ldots \right\}$$

obtained by letting $\boldsymbol{\beta}$ range over $B \subseteq \mathbb{R}^p, p \in \mathbb{N}$, where for each $t = 1, 2, \ldots$ and each $\boldsymbol{\beta} \in B$, $g_t(\cdot, \boldsymbol{\beta}) : \mathbb{R}^{\nu t} \to \mathbb{R}^+$ is measurable. We then refer to $g^T = \Pi_{t=1}^T g_t$ as the *quasi-likelihood*. Note that g_t has similar meaning as a conditional density, however for more details we refer interested readers to White (1994).

We define

$$L_T(\mathbf{x}^T, \boldsymbol{\beta}) \equiv T^{-1} \log\, g^T(\mathbf{x}^T, \boldsymbol{\beta}) = T^{-1} \sum_{t=1}^{T} \log\, g_t(\mathbf{x}^t, \boldsymbol{\beta}). \tag{2.10}$$

Let $\hat{\boldsymbol{\beta}}_T$ be a solution to the problem

$$\max_{\boldsymbol{\beta} \in B} L_T(\mathbf{x}^T, \boldsymbol{\beta}).$$

The function L_T is called the *quasi-log-likelihood function*, and a solution $\hat{\boldsymbol{\beta}}_T$ is called a *quasi-maximum likelihood estimator* (QMLE). For more details and the motivation to define such an estimator see White (1994). Given particular assumptions it can be shown that a QMLE $\hat{\boldsymbol{\beta}}_T$ exists for each T that maximizes the quasi-log-likelihood function with probability one.

We want to investigate statistical properties of the regression θ-quantile estimator, where $0 < \theta < 1$. Let Y_t denote dependent variables and let W_t denote vectors of explanatory variables of dimension k_t. Explanatory variables belong to the information set available at time t which is denoted as Ω_t. Then we specify a model for the conditional θ-quantile of dependent variables Y_t given explanatory variables W_t as a collection of sequences

$$f(\boldsymbol{\beta}) = \left\{ f_t(W_t, \boldsymbol{\beta}) : \mathbb{R}^{k_t} \times B \to \mathbb{R} \right\}$$

with the following assumption.

Assumption W0 :

(a) The function $f_t : \mathbb{R}^{k_t} \times B \to \mathbb{R}$ is such that for each $\boldsymbol{\beta} \in B$, a compact subset of $\mathbb{R}^p$, $f_t(\cdot, \boldsymbol{\beta})$ is measurable with respect to the information set Ω_t and $f_t(W_t, \cdot)$ is continuous in $B, t = 1, 2, \ldots$, for a given choice of explanatory variables $\{W_t\}$.

(b) There exists $\boldsymbol{\beta}^0$ in B such that

$$\theta = \mathsf{P}\big[Y_t \leq f_t(W_t, \boldsymbol{\beta}^0) | W_t\big], \ t = 1, 2, \ldots .$$

The vector $\boldsymbol{\beta}^0$ defined in the Assumption W0 (b) is the vector of true unknown parameters and $f_t(W_t, \boldsymbol{\beta}^0)$ is the conditional θ-quantile of Y_t given W_t. From now on we will denote $f_t(\mathbf{w}_t, \boldsymbol{\beta})$ only as $f_t(\boldsymbol{\beta})$ for simplicity.

Let the functions g_t in (2.10) be given by $g_t(\mathbf{x}^t, \boldsymbol{\beta}) = \exp \phi_t(y_t, f_t(\boldsymbol{\beta}))$, where $\phi_t : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is defined as

$$\phi_t(y, f) = -|y - f|\big[\theta \ I(y \geq f) + (1 - \theta) \ I(y < f)\big],$$

and $I(\cdot)$ is the indicator function for the specified event, $t = 1, 2, \ldots$.

Since

$$\sum_{t=1}^{T} \phi_t(y_t, f_t) = -\left\{ \sum_{t:\ y_t \geq f_t} \theta \mid y_t - f_t \mid + \sum_{t:\ y_t < f_t} (1 - \theta) \mid y_t - f_t \mid \right\},$$

we can write

$$\begin{aligned} L_T(\mathbf{x}^T, \boldsymbol{\beta}) \ &\equiv \ T^{-1} \log \, g^T(\mathbf{x}^T, \boldsymbol{\beta}) = T^{-1} \sum_{t=1}^{T} \log \, g_t(\mathbf{x}^t, \boldsymbol{\beta}) = T^{-1} \sum_{t=1}^{T} \phi_t\big(y_t, f_t(\boldsymbol{\beta})\big) \\ &= \ - \ T^{-1} \left\{ \sum_{t:\ y_t \geq f_t(\boldsymbol{\beta})} \theta \mid y_t - f_t(\boldsymbol{\beta}) \mid + \sum_{t:\ y_t < f_t(\boldsymbol{\beta})} (1 - \theta) \mid y_t - f_t(\boldsymbol{\beta}) \mid \right\}, \end{aligned}$$

which can be rewritten (in the same way as (2.6) can be rewritten as (2.7)) as

$$L_T(\mathbf{x}^T, \boldsymbol{\beta}) = -T^{-1} \sum_{t=1}^{T} \left\{ \Big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big)\Big] \big[y_t - f_t(\boldsymbol{\beta})\big] \right\}.$$

The QMLE $\hat{\boldsymbol{\beta}}_T$ is defined as a solution to the problem

$$\max_{\boldsymbol{\beta} \in B} L_T(\mathbf{x}^T, \boldsymbol{\beta}) = \max_{\boldsymbol{\beta} \in B} -T^{-1} \sum_{t=1}^{T} \left\{ \Big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big)\Big] \big[y_t - f_t(\boldsymbol{\beta})\big] \right\},$$

consequently, the QMLE $\hat{\boldsymbol{\beta}}_T$ is a solution to

$$\min_{\boldsymbol{\beta} \in B} T^{-1} \sum_{t=1}^{T} \left\{ \Big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big)\Big] \big[y_t - f_t(\boldsymbol{\beta})\big] \right\}. \tag{2.11}$$

Let

$$Q_T(\boldsymbol{\beta}) \equiv T^{-1} \sum_{t=1}^{T} q_t(\boldsymbol{\beta}),$$

where

$$q_t(\boldsymbol{\beta}) \equiv \Big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big)\Big]\big[y_t - f_t(\boldsymbol{\beta})\big].$$

Assumption W1 :

(a) For each $\boldsymbol{\beta}$ in B: $\mathsf{E}\big[q_t(\boldsymbol{\beta})\big]$ exists and is finite, $t = 1, 2, \ldots$;

(b) $\mathsf{E}\big[q_t(\cdot)\big]$ is continuous on $B, t = 1, 2, \ldots$; and

(c) $\big\{q_t(\boldsymbol{\beta})\big\}$ obeys the uniform law of large numbers.

Assumption W2 : $\mathsf{E}\big[Q_T(\boldsymbol{\beta}) - Q_T(\boldsymbol{\beta}^0)\big]$ is uniquely minimized at $\boldsymbol{\beta}^0$ for T sufficiently large, where $\boldsymbol{\beta}^0$ is the vector of true unknown parameters defined in the Assumption W0 (b).

Corollary W : (Corollary 5.12 of White (1994)) Assume that $(\Omega, \mathcal{F}, \mathsf{P})$ is a complete probability space and the observed data are a realization of a stochastic process on this space. Let Assumptions W0, W1 and W2 hold true. Then the QMLE estimator $\hat{\boldsymbol{\beta}}_T$ given as a solution to (2.11) satisfies

$$\hat{\boldsymbol{\beta}}_T \xrightarrow{\mathsf{P}} \boldsymbol{\beta}^0 \quad \text{as} \quad T \to \infty,$$

where $\boldsymbol{\beta}^0$ is the vector of true unknown parameters defined in the Assumption W0 (b) and the symbol $\xrightarrow{\mathsf{P}}$ denotes the convergence in probability.

Proof of the Theorem 1

Proof. We will verify that all conditions of the Corollary W are satisfied. The general assumptions and the Assumption W0 are satisfied according to Assumptions C0 and C1. It remains to check Assumptions W1 and W2. We have

$$Q_T(\boldsymbol{\beta}) \equiv T^{-1} \sum_{t=1}^{T} q_t(\boldsymbol{\beta}),$$

where

$$q_t(\boldsymbol{\beta}) \equiv \Big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big)\Big]\big[y_t - f_t(\boldsymbol{\beta})\big].$$

First, we need to show that $\mathsf{E}\big[q_t(\boldsymbol{\beta})\big]$ exists and is finite for every $\boldsymbol{\beta}$. This fact can be easily checked as follows (using triangle inequality and the definition $y_t \equiv f_t(\boldsymbol{\beta}^0) + \varepsilon_{t\theta}, t = 1, \ldots, T$)

$$\mathsf{E}\big[q_t(\boldsymbol{\beta})\big] < \mathsf{E}\big|y_t - f_t(\boldsymbol{\beta})\big| \leq \mathsf{E}|\varepsilon_{t\theta}| + \mathsf{E}\big|f_t(\boldsymbol{\beta})\big| + \mathsf{E}\big|f_t(\boldsymbol{\beta}^0)\big| < \infty,$$

by Assumptions C4 and C5. Moreover, because f_t is continuous in $\boldsymbol{\beta}$ by the Assumption C1, $q_t(\boldsymbol{\beta})$ is continuous (because the regression quantile objective function is continuous in $f_t(\boldsymbol{\beta})$), and hence its expected value (which we just showed

to be finite) will be also continuous. Therefore, Assumptions W1 (a) and (b) are satisfied. The Assumption C6 imply that the Assumption W1 (c) is also satisfied.

It remains to show that $\mathsf{E}\big[V_T(\boldsymbol{\beta})\big] \equiv \mathsf{E}\big[Q_T(\boldsymbol{\beta}) - Q_T(\boldsymbol{\beta}^0)\big]$ is uniquely minimized at $\boldsymbol{\beta}^0$ for T sufficiently large. Let

$$\nu_t(\boldsymbol{\beta}) \equiv q_t(\boldsymbol{\beta}) - q_t(\boldsymbol{\beta}^0).$$

Note that (using the fact that $y_t \equiv f_t(\boldsymbol{\beta}^0) + \varepsilon_{t\theta}, t = 1, \ldots, T$)

$$\begin{aligned} q_t(\boldsymbol{\beta}) &= \big[\theta - I\big(y_t < f_t(\boldsymbol{\beta})\big)\big]\big[y_t - f_t(\boldsymbol{\beta})\big] \\ &= \big[\theta - I\big(\varepsilon_{t\theta} < f_t(\boldsymbol{\beta}) - f_t(\boldsymbol{\beta}^0)\big)\big]\big[f_t(\boldsymbol{\beta}^0) + \varepsilon_{t\theta} - f_t(\boldsymbol{\beta})\big] \\ &= \big[\theta - I\big(\varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta})\big)\big]\big[\varepsilon_{t\theta} - \delta_t(\boldsymbol{\beta})\big], \end{aligned}$$

where

$$\delta_t(\boldsymbol{\beta}) \equiv f_t(\boldsymbol{\beta}) - f_t(\boldsymbol{\beta}^0).$$

Specially, it is obvious that $\delta_t(\boldsymbol{\beta}^0) = 0$.

Then

$$q_t(\boldsymbol{\beta}) = \begin{cases} [\theta - 1]\big[\varepsilon_{t\theta} - \delta_t(\boldsymbol{\beta})\big] & \text{if } \varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta}), \\ \theta\big[\varepsilon_{t\theta} - \delta_t(\boldsymbol{\beta})\big] & \text{if } \varepsilon_{t\theta} \geq \delta_t(\boldsymbol{\beta}). \end{cases}$$

Specially for $\boldsymbol{\beta}^0$,

$$q_t(\boldsymbol{\beta}^0) = \begin{cases} [\theta - 1]\varepsilon_{t\theta} & \text{if } \varepsilon_{t\theta} < 0, \\ \theta\varepsilon_{t\theta} & \text{if } \varepsilon_{t\theta} \geq 0. \end{cases}$$

Consider the cases:

- if $\varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta})$ and $\varepsilon_{t\theta} < 0$ then
  $\nu_t(\boldsymbol{\beta}) = [\theta - 1]\big[\varepsilon_{t\theta} - \delta_t(\boldsymbol{\beta})\big] - [\theta - 1]\varepsilon_{t\theta} = (1 - \theta)\delta_t(\boldsymbol{\beta})$,
- if $\varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta})$ and $\varepsilon_{t\theta} \geq 0$ then
  $\nu_t(\boldsymbol{\beta}) = [\theta - 1]\big[\varepsilon_{t\theta} - \delta_t(\boldsymbol{\beta})\big] - \theta\varepsilon_{t\theta} = (1 - \theta)\delta_t(\boldsymbol{\beta}) - \varepsilon_{t\theta}$,
- if $\varepsilon_{t\theta} \geq \delta_t(\boldsymbol{\beta})$ and $\varepsilon_{t\theta} < 0$ then
  $\nu_t(\boldsymbol{\beta}) = \theta\big[\varepsilon_{t\theta} - \delta_t(\boldsymbol{\beta})\big] - [\theta - 1]\varepsilon_{t\theta} = \varepsilon_{t\theta} - \theta\delta_t(\boldsymbol{\beta})$,
- if $\varepsilon_{t\theta} \geq \delta_t(\boldsymbol{\beta})$ and $\varepsilon_{t\theta} \geq 0$ then
  $\nu_t(\boldsymbol{\beta}) = \theta\big[\varepsilon_{t\theta} - \delta_t(\boldsymbol{\beta})\big] - \theta\varepsilon_{t\theta} = -\theta\delta_t(\boldsymbol{\beta})$.

Therefore,

$$\nu_t(\boldsymbol{\beta}) = \begin{cases} (1 - \theta)\delta_t(\boldsymbol{\beta}) & \text{if } \varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta}) \text{ and } \varepsilon_{t\theta} < 0, \\ (1 - \theta)\delta_t(\boldsymbol{\beta}) - \varepsilon_{t\theta} & \text{if } \varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta}) \text{ and } \varepsilon_{t\theta} \geq 0, \\ \varepsilon_{t\theta} - \theta\delta_t(\boldsymbol{\beta}) & \text{if } \varepsilon_{t\theta} \geq \delta_t(\boldsymbol{\beta}) \text{ and } \varepsilon_{t\theta} < 0, \\ -\theta\delta_t(\boldsymbol{\beta}) & \text{if } \varepsilon_{t\theta} \geq \delta_t(\boldsymbol{\beta}) \text{ and } \varepsilon_{t\theta} \geq 0. \end{cases}$$

Now, we will show that

$$\begin{aligned} \mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big] \quad &= \quad I\big(\delta_t(\boldsymbol{\beta}) < 0\big) \int_{-|\delta_t(\boldsymbol{\beta})|}^{0} \big(\lambda + |\delta_t(\boldsymbol{\beta})|\big) h_t(\lambda|\Omega_t)\, d\lambda \\ &+ \quad I\big(\delta_t(\boldsymbol{\beta}) > 0\big) \int_{0}^{|\delta_t(\boldsymbol{\beta})|} \big(|\delta_t(\boldsymbol{\beta})| - \lambda\big) h_t(\lambda|\Omega_t)\, d\lambda. \end{aligned}$$

We need to evaluate $\mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big]$. According to the Assumption C2 the error terms $\varepsilon_{t\theta}$, conditional on all of the past information Ω_t, form a stationary process, with continuous conditional density $h_t(\varepsilon|\Omega_t)$. That implies

$$\int_{-\infty}^{\infty} h_t(\lambda|\Omega_t)\ d\lambda = 1. \tag{2.12}$$

Consider two possibilities[2]: $\delta_t(\boldsymbol{\beta}) < 0$ and $\delta_t(\boldsymbol{\beta}) > 0$.

If $\delta_t(\boldsymbol{\beta}) < 0$ then

$$\nu_t(\boldsymbol{\beta}) = \begin{cases} (1-\theta)\delta_t(\boldsymbol{\beta}) & \text{if } \varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta}), \\ \varepsilon_{t\theta} - \theta\delta_t(\boldsymbol{\beta}) & \text{if } \delta_t(\boldsymbol{\beta}) \le \varepsilon_{t\theta} < 0, \\ -\theta\delta_t(\boldsymbol{\beta}) & \text{if } 0 \le \varepsilon_{t\theta}. \end{cases}$$

Consequently,

$$\begin{aligned} \mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big] &= \int_{-\infty}^{\delta_t(\boldsymbol{\beta})} (1-\theta)\delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda + \int_{\delta_t(\boldsymbol{\beta})}^{0} \big(\lambda - \theta\delta_t(\boldsymbol{\beta})\big)\ h_t(\lambda|\Omega_t)\ d\lambda \\ &\qquad - \int_{0}^{\infty} \theta\delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda \\ &= \int_{-\infty}^{\delta_t(\boldsymbol{\beta})} \delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda\ +\ \int_{\delta_t(\boldsymbol{\beta})}^{0} \lambda\ h_t(\lambda|\Omega_t)\ d\lambda \\ &\qquad - \int_{-\infty}^{\infty} \theta\delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda. \end{aligned} \tag{2.13}$$

The Assumption C2 and formulation of our model (2.8), that is

$$y_t = f(y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1; \boldsymbol{\beta}^0) + \varepsilon_{t\theta} \equiv f_t(\boldsymbol{\beta}^0) + \varepsilon_{t\theta}, \qquad t = 1, \ldots, T,$$

$$\Omega_t = \big[y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1, f_1(\boldsymbol{\beta}^0)\big],$$

$$Quant_\theta(\varepsilon_{t\theta}|\Omega_t) = 0,$$

imply, together with (2.12), that

$$\int_{-\infty}^{\infty} \theta\delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda = \theta\delta_t(\boldsymbol{\beta}) \int_{-\infty}^{\infty} h_t(\lambda|\Omega_t)\ d\lambda = \theta\delta_t(\boldsymbol{\beta}), \tag{2.14}$$

and also

$$\begin{aligned} \int_{-\infty}^{\delta_t(\boldsymbol{\beta})} \delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda &= \int_{-\infty}^{\infty} \delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda\ -\ \int_{\delta_t(\boldsymbol{\beta})}^{\infty} \delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda \\ &= \delta_t(\boldsymbol{\beta})\ -\ \int_{\delta_t(\boldsymbol{\beta})}^{\infty} \delta_t(\boldsymbol{\beta})\ h_t(\lambda|\Omega_t)\ d\lambda. \end{aligned} \tag{2.15}$$

[2]Note that if $\delta_t(\boldsymbol{\beta}) = 0$ then

$$\nu_t(\boldsymbol{\beta}) = \begin{cases} (1-\theta)\delta_t(\boldsymbol{\beta}) = 0 & \text{if } \varepsilon_{t\theta} < 0, \\ -\theta\delta_t(\boldsymbol{\beta}) = 0 & \text{if } \varepsilon_{t\theta} \ge 0, \end{cases}$$

that is $\nu_t(\boldsymbol{\beta}) = 0$. As a result, $I\big(\delta_t(\boldsymbol{\beta}) = 0\big) \cdot \mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big] = 0$.

If we insert (2.14) and (2.15) into (2.13), we get (remember that now we consider the case $\delta_t(\boldsymbol{\beta}) < 0$)

$$\begin{aligned}
&\mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big] \;= \\
&= \delta_t(\boldsymbol{\beta}) \;-\; \int_{\delta_t(\boldsymbol{\beta})}^{\infty} \delta_t(\boldsymbol{\beta})\; h_t(\lambda|\Omega_t)\; d\lambda \;+\; \int_{\delta_t(\boldsymbol{\beta})}^{0} \lambda\; h_t(\lambda|\Omega_t)\; d\lambda \;-\; \theta\delta_t(\boldsymbol{\beta}) \\
&= \int_{-|\delta_t(\boldsymbol{\beta})|}^{0} \big(\lambda + |\delta_t(\boldsymbol{\beta})|\big)\; h_t(\lambda|\Omega_t)\; d\lambda \;-\; \delta_t(\boldsymbol{\beta}) \int_0^{\infty} h_t(\lambda|\Omega_t) \;+\; (1-\theta)\delta_t(\boldsymbol{\beta}),
\end{aligned}$$

The equation

$$Quant_\theta(\varepsilon_{t\theta}|\Omega_t) = 0,$$

means that

$$\int_{-\infty}^{0} h_t(\lambda|\Omega_t)\; d\lambda = \theta, \quad \text{resp.} \int_0^{\infty} h_t(\lambda|\Omega_t)\; d\lambda = 1-\theta. \tag{2.16}$$

Therefore,

$$\begin{aligned}
\mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big] \;&=\; \int_{-|\delta_t(\boldsymbol{\beta})|}^{0} \big(\lambda + |\delta_t(\boldsymbol{\beta})|\big)\; h_t(\lambda|\Omega_t)\; d\lambda \;-\delta_t(\boldsymbol{\beta})(1-\theta) \;+(1-\theta)\delta_t(\boldsymbol{\beta}) \\
&=\; \int_{-|\delta_t(\boldsymbol{\beta})|}^{0} \big(\lambda + |\delta_t(\boldsymbol{\beta})|\big)\; h_t(\lambda|\Omega_t)\; d\lambda.
\end{aligned} \tag{2.17}$$

If $\delta_t(\boldsymbol{\beta}) > 0$ then

$$\nu_t(\boldsymbol{\beta}) = \begin{cases} (1-\theta)\delta_t(\boldsymbol{\beta}) & \text{if } \varepsilon_{t\theta} < 0, \\ (1-\theta)\delta_t(\boldsymbol{\beta}) - \varepsilon_{t\theta} & \text{if } 0 \le \varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta}), \\ -\theta\delta_t(\boldsymbol{\beta}) & \text{if } \delta_t(\boldsymbol{\beta}) \le \varepsilon_{t\theta}. \end{cases}$$

Applying (2.12), (2.8) and (2.16) we analogously get

$$\begin{aligned}
\mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big] \;&=\; \int_{-\infty}^{0} (1-\theta)\delta_t(\boldsymbol{\beta})\; h_t(\lambda|\Omega_t)\; d\lambda \\
&\quad + \int_0^{\delta_t(\boldsymbol{\beta})} \big((1-\theta)\delta_t(\boldsymbol{\beta}) - \lambda\big) h_t(\lambda|\Omega_t) d\lambda - \int_{\delta_t(\boldsymbol{\beta})}^{\infty} \theta\delta_t(\boldsymbol{\beta}) h_t(\lambda|\Omega_t) d\lambda \\
&=\; \int_{-\infty}^{0} \delta_t(\boldsymbol{\beta})\; h_t(\lambda|\Omega_t)\; d\lambda \;+\; \int_0^{\delta_t(\boldsymbol{\beta})} \big(\delta_t(\boldsymbol{\beta}) - \lambda\big)\; h_t(\lambda|\Omega_t)\; d\lambda \\
&\qquad\qquad - \int_{-\infty}^{\infty} \theta\delta_t(\boldsymbol{\beta})\; h_t(\lambda|\Omega_t)\; d\lambda \\
&=\; \delta_t(\boldsymbol{\beta}) \int_{-\infty}^{0} h_t(\lambda|\Omega_t)\; d\lambda \;+\; \int_0^{|\delta_t(\boldsymbol{\beta})|} \big(|\delta_t(\boldsymbol{\beta})| - \lambda\big)\; h_t(\lambda|\Omega_t)\; d\lambda \\
&\qquad\qquad - \theta\delta_t(\boldsymbol{\beta}) \int_{-\infty}^{\infty} h_t(\lambda|\Omega_t)\; d\lambda \\
&=\; \theta\delta_t(\boldsymbol{\beta}) \;+\; \int_0^{|\delta_t(\boldsymbol{\beta})|} \big(|\delta_t(\boldsymbol{\beta})| - \lambda\big)\; h_t(\lambda|\Omega_t)\; d\lambda \;-\; \theta\delta_t(\boldsymbol{\beta}) \\
&=\; \int_0^{|\delta_t(\boldsymbol{\beta})|} \big(|\delta_t(\boldsymbol{\beta})| - \lambda\big)\; h_t(\lambda|\Omega_t)\; d\lambda.
\end{aligned} \tag{2.18}$$

Finally, if we put (2.17) and (2.18) together, we get

$$\begin{aligned}\mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big] \quad &= \quad I\big(\delta_t(\boldsymbol{\beta})<0\big)\int_{-|\delta_t(\boldsymbol{\beta})|}^{0}\big(\lambda+|\delta_t(\boldsymbol{\beta})|\big)h_t(\lambda|\Omega_t)\;d\lambda \\ &+ \quad I\big(\delta_t(\boldsymbol{\beta})>0\big)\int_{0}^{|\delta_t(\boldsymbol{\beta})|}\big(|\delta_t(\boldsymbol{\beta})|-\lambda\big)h_t(\lambda|\Omega_t)\;d\lambda.\end{aligned}$$

The continuity of $h_t(\cdot|\Omega_t)$ (the Assumption C2) and the Assumption C3 imply that there exists $h_1>0$ such that $h_t(\lambda|\Omega_t)>h_1$ whenever $|\lambda|<h_1$. Hence, for any $0<\tau<h_1$ the following holds

$$\begin{aligned}\mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big] \;&\geq\; I\big(\delta_t(\boldsymbol{\beta})<-\tau\big)\int_{-\tau}^{0}(\lambda+\tau)\,h_1 d\lambda + I\big(\delta_t(\boldsymbol{\beta})>\tau\big)\int_{0}^{\tau}(\tau-\lambda)\,h_1 d\lambda \\ &=\; \frac{1}{2}\,\tau^2\;h_1\;I(|\delta_t(\boldsymbol{\beta})|>\tau).\end{aligned}$$

Taking the unconditional expectations we get

$$\begin{aligned}\mathsf{E}\big[\nu_t(\boldsymbol{\beta})\big] \;&=\; \mathsf{E}\Big[\mathsf{E}\big[\nu_t(\boldsymbol{\beta})|\Omega_t\big]\Big] \\ &\geq\; \frac{1}{2}\,\tau^2\;h_1\;\mathsf{P}\big[|\delta_t(\boldsymbol{\beta})|>\tau\big]=\frac{1}{2}\,\tau^2\;h_1\;\mathsf{P}\Big[\big|f_t(\boldsymbol{\beta})-f_t(\boldsymbol{\beta}^0)\big|>\tau\Big].\end{aligned}$$

As a consequence,

$$\begin{aligned}\mathsf{E}\big[V_T(\boldsymbol{\beta})\big] \;&\equiv\; \mathsf{E}\big[Q_T(\boldsymbol{\beta})-Q_T(\boldsymbol{\beta}^0)\big] \;=\; \mathsf{E}\left[T^{-1}\sum_{t=1}^{T}q_t(\boldsymbol{\beta})-T^{-1}\sum_{t=1}^{T}q_t(\boldsymbol{\beta}^0)\right] \\ &=\; \mathsf{E}\left[T^{-1}\sum_{t=1}^{T}\big(q_t(\boldsymbol{\beta})-q_t(\boldsymbol{\beta}^0)\big)\right] \;=\; \mathsf{E}\left[T^{-1}\sum_{t=1}^{T}\nu_t(\boldsymbol{\beta})\right] \\ &=\; T^{-1}\sum_{t=1}^{T}\mathsf{E}\big[\nu_t(\boldsymbol{\beta})\big] \;\geq\; \frac{1}{2}\,\tau^2\;h_1\;T^{-1}\sum_{t=1}^{T}\mathsf{P}\Big[\big|f_t(\boldsymbol{\beta})-f_t(\boldsymbol{\beta}^0)\big|>\tau\Big],\end{aligned}$$

which is greater than 0 by the Assumption C7 if $\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|\geq\xi$. This means that also the Assumption W2 is satisfied.

We verified the conditions of the Corollary W, thus the proof of the Theorem 1 is complete. □

Theorem 2 (Asymptotic normality). Consider the model (2.8). Let the conditions of the Theorem 1 and the following assumptions be satisfied:

AN1. $f_t(\boldsymbol{\beta})$ is differentiable in B, $\boldsymbol{\beta}^0$ is an interior point of B and for all $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in a neighborhood υ_0 of $\boldsymbol{\beta}^0$, such that $\|\boldsymbol{\beta}-\boldsymbol{\gamma}\|\leq d$ for d sufficiently small and for all t:

(a) $\|\nabla f_t(\boldsymbol{\beta})\|\leq F(\Omega_t)$, where $F(\Omega_t)$ is some (possibly) stochastic function of variables that belong to the information set and $\mathsf{E}\big[F(\Omega_t)^3\big]\leq F_0<\infty$, for some constant F_0.

(b) $\|\nabla f_t(\boldsymbol{\beta}) - \nabla f_t(\boldsymbol{\gamma})\| \leq M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma}) = O(\|\boldsymbol{\beta} - \boldsymbol{\gamma}\|)$, where $M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})$ is some function such that $\mathsf{E}[M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})]^2 \leq M_0\|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$ and $\mathsf{E}[M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})F(\Omega_t)] \leq M_1\|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$ for some constants M_0 and M_1.

AN2. (a) $h_t(\varepsilon|\Omega_t) \leq N < \infty \ \forall t$, for some constant N.

(b) $h_t(\varepsilon|\Omega_t)$ satisfies the Lipschitz condition $|h_t(\lambda_1|\Omega_t) - h_t(\lambda_2|\Omega_t)| \leq L|\lambda_1 - \lambda_2|$ for some constant $L < \infty \ \forall t$.

AN3. The matrices $\mathrm{A}_T \equiv \mathsf{E}\big[T^{-1}\theta(1-\theta)\sum_{t=1}^{T} \nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\big]$ and $\mathrm{D}_T \equiv \mathsf{E}\big[T^{-1}\sum_{t=1}^{T} h_t(0|\Omega_t)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\big]$ have the smallest eigenvalues bounded below by a positive constant for T sufficiently large.

AN4. The sequence $\big\{T^{-1/2}\sum_{t=1}^{T}\big[\theta - I\big(y_t < f_t(\boldsymbol{\beta}^0)\big)\big]\nabla' f_t(\boldsymbol{\beta}^0)\big\}$ obeys the central limit theorem.

Then

$$\sqrt{T}\mathrm{A}_T^{-1/2}\mathrm{D}_T(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}^0) \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, \mathrm{I}),$$

where

$$\mathrm{A}_T \equiv \mathsf{E}\left[T^{-1}\theta(1-\theta)\sum_{t=1}^{T} \nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\right],$$

$$\mathrm{D}_T \equiv \mathsf{E}\left[T^{-1}\sum_{t=1}^{T} h_t(0|\Omega_t)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\right],$$

$\hat{\boldsymbol{\beta}}_T$ is computed as in the Theorem 1 and the symbol $\xrightarrow{\mathrm{d}}$ denotes the convergence in distribution.

Theorem 3 (Variance-covariance matrix estimation). Consider the model (2.8). Let the conditions of the Theorem 2 (and also the Theorem 1) and the following assumptions be satisfied:

VC1. $\hat{c}_T/c_T \xrightarrow{\mathsf{P}} 1$, where the nonstochastic positive sequence c_T satisfies $c_T = o(1)$ and $c_T^{-1} = o(T^{1/2})$.

VC2. $\mathsf{E}\big[|F(\Omega_t)|^4\big] \leq F_1 < \infty$ for all t and for some constant F_1, where $F(\Omega_t)$ has been defined in the Assumption AN1(a).

VC3. $T^{-1}\theta(1-\theta)\sum_{t=1}^{T} \nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0) - \mathrm{A}_T \xrightarrow{\mathsf{P}} 0$ and $T^{-1}\sum_{t=1}^{T} h_t(0|\Omega_t)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0) - \mathrm{D}_T \xrightarrow{\mathsf{P}} 0$.

Then $\hat{\mathrm{A}}_T - \mathrm{A}_T \xrightarrow{\mathsf{P}} \mathbf{0}$ and $\hat{\mathrm{D}}_T - \mathrm{D}_T \xrightarrow{\mathsf{P}} \mathbf{0}$, where

$$\hat{\mathrm{A}}_T \equiv T^{-1}\theta(1-\theta)\nabla' f(\hat{\boldsymbol{\beta}}_T)\nabla f(\hat{\boldsymbol{\beta}}_T),$$

$$\hat{\mathrm{D}}_T \equiv (2T\hat{c}_T)^{-1}\sum_{t=1}^{T} I\left(\big|y_t - f_t(\hat{\boldsymbol{\beta}}_T)\big| < \hat{c}_T\right)\nabla' f_t(\hat{\boldsymbol{\beta}}_T)\nabla f_t(\hat{\boldsymbol{\beta}}_T),$$

A_T and D_T have been defined in the Theorem 2, $\hat{\boldsymbol{\beta}}_T$ is defined as in the Theorem 1, $\hat{c}_T$ is a bandwidth defined in the Assumption VC1 and the symbol $\xrightarrow{\mathsf{P}}$ denotes the convergence in probability.

The proofs of Theorems 2 and 3 can be found in Engle and Manganelli (2004). The proof of the Theorem 2 is based on the Theorem 3 in Huber (1967) and the work of Weiss (1991).

2.3 Evaluating VaR models

There exist various methods how to assess the performance of different models used for predicting VaR. One of the most popular (and also the simplest) is to check whether the correct fraction of VaR exceeding is achieved (i.e. to check how often the losses have exceeded the estimated VaR).

Consider we have N out-of-sample observations of returns y_t and N corresponding one day θ VaR estimates $\widehat{\mathrm{VaR}}_t$, $t = 1, \ldots, N$. If all assumptions of the model which was used for predicting VaR are satisfied, then one expects that

$$\mathsf{P}[y_t < -\widehat{\mathrm{VaR}}_t] = \theta, \quad t = 1, \ldots, N.$$

Define

$$H_t \equiv I(y_t < -\widehat{\mathrm{VaR}}_t), \quad t = 1, \ldots, N.$$

Correct fraction of VaR exceeding means that the number of VaR exceeding (i.e. the sum of hits H_t) divided by N, that is

$$\frac{1}{N} \sum_{t=1}^{N} H_t,$$

is close to θ. This is what we expect and it is also the simplest heuristics how to assess the performance of a quantile model. However, that is not the only requirement that every good quantile estimate should satisfy. It is desirable that a violation today has no influence on the probability of violation tomorrow. VaR forecasts that do not consider time-varying volatility may be correct on average, but will produce violation clusters (see Christoffersen (1998)). Therefore, numerous more sophisticated method for evaluating VaR models have been proposed.

In the work of Christoffersen (1998) several tests are suggested for evaluating the accuracy of out-of-sample interval forecasts. Christoffersen (1998) defines the sequence of VaR forecasts efficient with respect to Ω_{t-1} if

$$\mathsf{E}[H_t | \Omega_{t-1}] = \theta. \tag{2.19}$$

If we apply iterated expectations, then equation (2.19) implies that H_t is uncorrelated with any function of a variable in the information set available at time $t-1$. Equation (2.19) also implies the correct unconditional coverage of the interval forecast (that is the correct fraction of VaR exceeding). Moreover, if equation

(2.19) holds, then VaR violations will occur with the correct conditional probability. By specifying Ω_{t-1} to include at least $\{H_1, H_2, \ldots, H_{t-1}\}$, it can be shown that equation (2.19) implies that $H_t|\Omega_t$ are i.i.d. random variables. For more details see Christoffersen (1998) or Kuester et al. (2006), where the test of unconditional (resp. conditional) coverage and the test of independence are described. These tests can reveal particular inadequacies of a model since they test specific implications of equation (2.19).

Dynamic quantile test

We will use the dynamic quantile test (DQ test) proposed by Engle and Manganelli (2004) which has various alternatives and can be used to examine different forms of misspecifications. The main idea is to regress $H_t - \theta$ on arbitrarily chosen explanatory variables in Ω_{t-1}. Generally, we have

$$H_t - \theta = \alpha_0 + \sum_{i=1}^{p+1} \alpha_i x_{t,i} + u_t, \quad t = 1, \ldots, N, \tag{2.20}$$

where regressors $x_{t,i},\ i = 1, \ldots, p+1$ belong to the information set Ω_{t-1} for all $t = 1, \ldots, N$. Under the null hypothesis (i.e. equation (2.19)) these regressors should have no explanatory power, that is $\alpha_i = 0,\ i = 0, \ldots, p+1$. Any function of the past information set which is suspected of being informative can be incorporated as an explanatory variable in (2.20), for instance one can choose

$$H_t - \theta = \alpha_0 + \sum_{i=1}^{p} \alpha_i (H_{t-i} - \theta) + \alpha_{p+1} \widehat{\mathrm{VaR}}_t + u_t. \tag{2.21}$$

We could also use annual dummy variables as regressors to check whether the correct fraction of VaR exceeding is satisfied each year.

In the matrix notation, we have

$$\boldsymbol{H} - \theta \mathbf{1} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{u}, \quad u_t = \begin{cases} -\theta & \text{with probability } 1-\theta, \\ 1-\theta & \text{with probability } \theta, \end{cases}$$

where $\boldsymbol{H} = (H_1, \ldots, H_N)'$ is the vector of hits, $\mathbf{1} = (1, \ldots, 1)'$ is the vector of ones, $\boldsymbol{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_N')'$ is a matrix, its rows $\mathbf{x}_t'$, $t = 1, \ldots, N$ are $(p+2)$-vectors measurable Ω_{t-1}, and $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p+1})'$. Under the null hypothesis $\boldsymbol{\alpha} = \mathbf{0}$. Engle and Manganelli (2004) deduce that under some additional assumptions the test statistic

$$\frac{1}{N} \frac{(\boldsymbol{H} - \theta\mathbf{1})' \boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' (\boldsymbol{H} - \theta\mathbf{1})}{\theta(1-\theta)}$$

has asymptotically the χ^2_{p+2}-distribution. If one use the first $m < p+2$ lagged hits as regressors in the matrix $\boldsymbol{X}$ then the vector $\boldsymbol{H}$ and the matrix $\boldsymbol{X}$ contain only rows $m+1, \ldots, N$. Note that Engle and Manganelli (2004) derive the in-sample dynamic quantile test as well, we refer interested readers to the original paper.

The out-of-sample DQ test does not depend on the model used for estimating VaR. To implement it, one just needs a sequence of VaR estimates and the corresponding sequence of returns. In the empirical analysis below, we use a constant, the VaR forecast and the first four lagged hits (that means we set $p = 4$ in (2.21)) as the explanatory variables contained in the regressor matrix $\boldsymbol{X}$.

2.4 Empirical results

We examined the forecasting performance of CAViaR models. We chose four American indices, namely Standard & Poor's 500 (S&P), NASDAQ Composite index (NASDAQ), Dow Jones Composite index (DJ) and New York Stock Exchange Composite index (NYSE). The data[3] consist of daily closing prices p_t from February 1, 1984, to February 1, 2008 (for all indices). Graphs of daily prices for all four indices are plotted in Figure 2.1.

We computed the daily returns scaled as

$$y_t = 100\ (\ln p_t - \ln p_{t-1}),$$

thus we have a total of 6 054 observations of daily returns. Graphs of daily returns for all four indices are plotted in Figure 2.2. We can clearly recognize the market crash in October 1987 and the period of high volatility in the years from 1998 to 2004. In Table 2.1 some basic statistics are presented. Notice that for all four indices the distribution of returns is negatively skewed and has significantly high kurtosis which is typical for daily financial data.

	Mean	Min	Max	Std. Dev.	Skewness	Kurtosis
S&P 500	0,035	-22,900	8,709	1,046	-1,933	45,478
NASDAQ	0,036	-12,043	13,255	1,360	-0,241	11,684
DJ	0,036	-22,475	8,419	0,991	-2,201	51,480
NYSE	0,037	-21,286	8,622	0,949	-2,165	50,056

Table 2.1: Summary statistics for daily returns of indices.

We used the first 5 054 observations to estimate the model and the remaining 1 000 observations for out-of-sample testing. We chose $\theta = 1\,\%$ and $5\,\%$ and estimated the one day VaR using the four CAViaR specifications suggested by Engle and Manganelli (2004), that is the symmetric absolute value (2.3) denoted as SAV, the asymmetric slope (2.4) denoted as AS, the indirect GARCH (2.5) and the adaptive (2.2). For the adaptive model (2.2), we set $\kappa = 10$. Note that we used larger sample than Engle and Manganelli (2004), therefore we can compare our findings with their results.

We need to compute $\hat{\boldsymbol{\beta}}_T$ which means to solve the minimization problem described by (2.9). Recall that $f_t(\boldsymbol{\beta})$ is defined to be the time t θ-quantile of the distribution

[3]The data were downloaded from http://finance.yahoo.com.

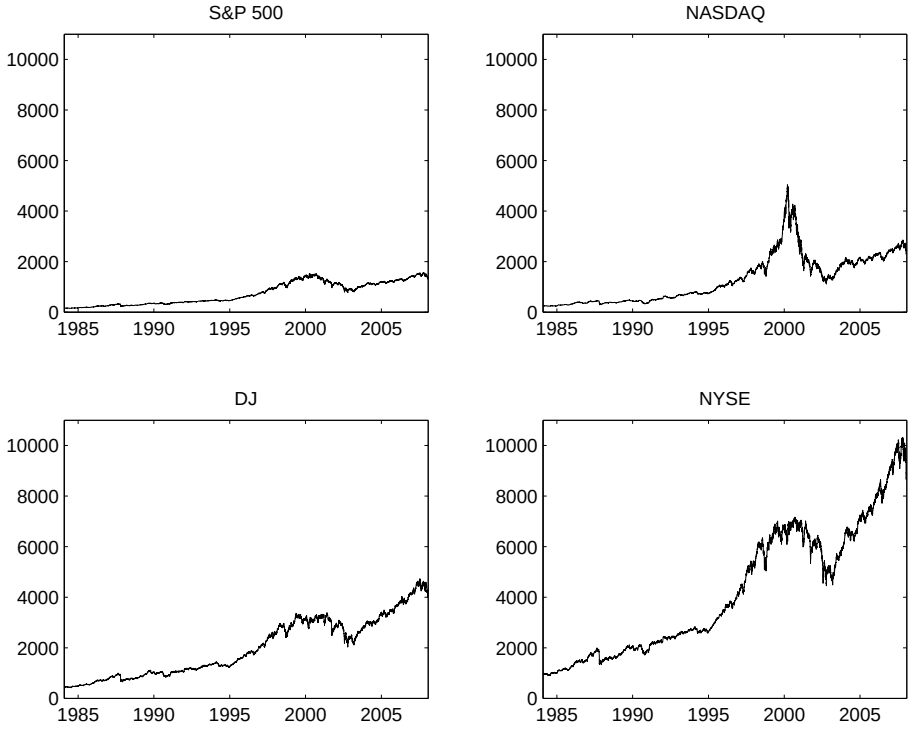


Figure 2.1: Daily prices of indices – all four graphs have the same scale on the y-axis.

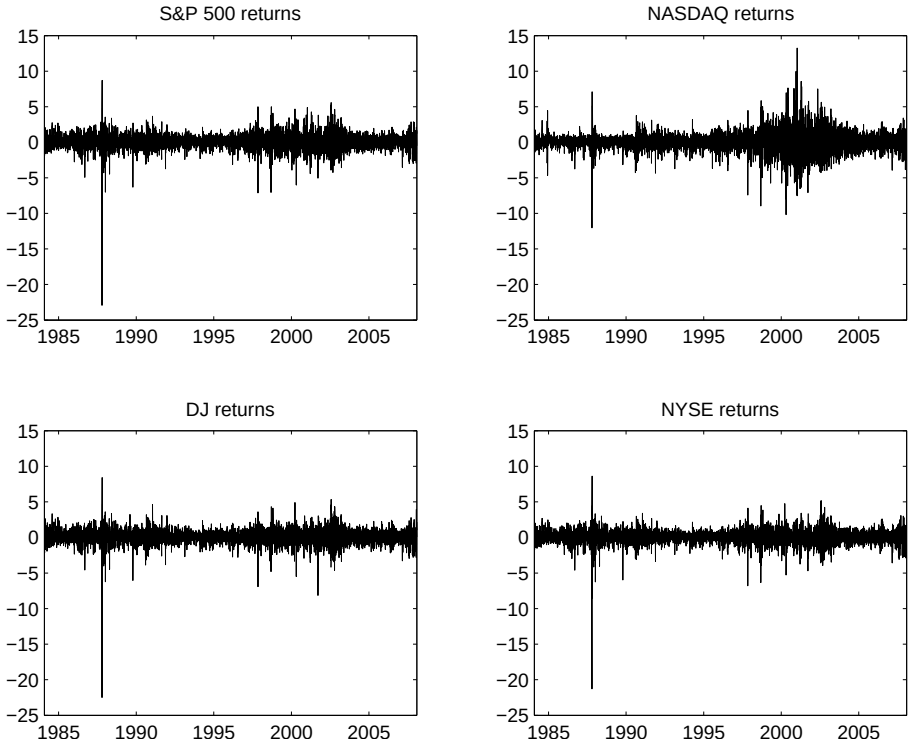


Figure 2.2: Daily returns of indices.

of portfolio returns formed at time $t-1$ and the one day VaR prediction for time t, denoted as $\text{VaR}_t(\boldsymbol{\beta})$, is equal to the time t negative θ-quantile of the distribution of portfolio returns formed at time $t-1$. Hence, in (2.9) we set $f_t(\boldsymbol{\beta}) = -\text{VaR}_t(\boldsymbol{\beta})$ and use the four CAViaR specifications mentioned above to compute $\text{VaR}_t(\boldsymbol{\beta})$. Consequently, the function f_t in (2.9) is not linear in the parameters $\boldsymbol{\beta}$ and is defined recursively. Moreover, the regression quantile function (RQ function), that is the objective function in (2.9), may have multiple local minima, therefore it is very difficult to find the exact solution to this minimization problem. We use the same implementation as Engle and Manganelli (2004)[4]. Now, we will describe the procedure.

To compute the VaR series with the CAViaR models, $\text{VaR}_1(\boldsymbol{\beta})$ was initialized to be the negative empirical θ-quantile of the first 300 observations (in other words, $f_1(\boldsymbol{\beta})$ was initialized to be the empirical θ-quantile of the first 300 observations). The loops to compute the recursive RQ functions for particular choice of parameters $\boldsymbol{\beta}$ were coded in C. All of the other computations were done in MATLAB R2007b, using the functions *fminsearch* and *fminunc* as optimization algorithms. Both of these function perform multidimensional unconstrained nonlinear minimization, that means they attempt to find a minimum of a scalar function of several variables, starting at an initial estimate. Function *fminsearch* uses a derivative-free method, namely the Nelder-Mead simplex method. Function *fminunc* uses the Quasi-Newton method and the minimized function must be continuous. Both functions may give only local solutions.

Engle and Manganelli (2004) suggest to use the following procedure which attempts to deal with the problem of local minima by using multiple starting points and combining both of the functions mentioned above. At the beginning, n vectors are generated using a uniform random number generator between 0 and 1. For each of these vectors the RQ function is computed and the m vectors that produced the lowest RQ criterion are selected as starting values for the optimization routine. Particularly, the parameters n and m are set $n = [10^4, 10^5, 10^4, 10^4]$ and $m = [10, 15, 10, 5]$ for the symmetric absolute value, asymmetric slope, indirect GARCH, and adaptive models. For each of these starting values, first the simplex algorithm is run. Then the optimal parameters from the simplex algorithm are used as starting points in the quasi-Newton algorithm and the new optimal parameters are used as the new starting conditions for the simplex. This procedure is repeated until the convergence criterion is satisfied. Tolerance levels for the function and the parameters values were set to 10^{-10}. Finally, the vector that produced the lowest RQ criterion is selected.

Tables 2.2, 2.3, 2.4 and 2.5 present the values of the estimated parameters, the corresponding standard errors and one-sided p-values[5], the values of the RQ function, the percentage of the times the VaR is exceeded (in-sample and out-of-sample),

[4]We are grateful to Simone Manganelli for providing his CAViaR programs.

[5]Note that the standard errors were computed using the variance-covariance matrix estimator (as described in Theorem 3). The one-sided p-values were computed using the asymptotic normality of $\hat{\boldsymbol{\beta}}$ (as described in Theorem 2) and the computed standard errors.

and the p-value of the out-of-sample DQ test for the case we used the first 5 054 observations to estimate the model and the remaining 1 000 observations for out-of-sample testing.[6]

In Tables 2.2, 2.3, 2.4 and 2.5 we can notice that the coefficients of the autoregressive terms (that is β_2) are invariably significant which corresponds to the fact that volatilities cluster in the tails as well and agrees with the results in Engle and Manganelli (2004). In the asymmetric slope model the coefficients of the positive part of lagged returns (that is β_3) are mostly not significantly different from zero (except for the NASDAQ), and on the contrary the coefficients of the negative part of lagged returns (that is β_4) are all significant except for NYSE 1% VaR. This fact reflects the asymmetry in the influence of lagged returns.

Notice that the percentage of in-sample hits for all four models and all four indices is very close to the expected 1 %, resp. 5 %. As Engle and Manganelli (2004) pointed out, the objective RQ function is directly designed to achieve this result.

In the case of 1% VaR the symmetric absolute value, the asymmetric slope and the indirect GARCH models overestimate VaR forecasts which results in poor out-of-sample performance, in particular hits out-of-sample are well below 1 %. This is not true for the adaptive model which produces accurate percentage of hits out-of-sample, especially for DJ and NYSE (1 %). In the case of 5% VaR all models slightly underestimate VaR forecasts (particularly the symmetric absolute value model), however the performance is much better than for 1% VaR.

Looking at all indices individually in the case of 1% VaR, we can see that for S&P the symmetric absolute value and the asymmetric slope models are rejected by the DQ test while the indirect GARCH and the adaptive models perform better. For NASDAQ and DJ the adaptive and the asymmetric slope models are better than the remaining models (the adaptive model produces precise percentage of out-of-sample hits). For NYSE the adaptive model gives the best results as well. To summarize, in the case of 1% VaR the adaptive model performs the best for all indices which differs from the results in Engle and Manganelli (2004).

In the case of 5% VaR, we can see that for S&P the adaptive and the asymmetric slope models perform better than the remaining models (the adaptive model produces precise percentage of out-of-sample hits). For NASDAQ the asymmetric slope model is rejected by the DQ test whereas the indirect GARCH model gives the best results. For DJ the asymmetric slope model gives better results than the other models producing rather accurate percentage of hits out-of-sample (5.1 %). For NYSE the adaptive model performs the best.

It is important to note, that the length of the out-of-sample period is 1 000 trading days. This approximately corresponds to four calendar years. It is probable

[6]Tables for the case when the first 5 554 observations were used to estimate the model and the remaining 500 observations for out-of-sample testing are in the Appendix (Tables A.1, A.2, A.3 and A.4).

that financial institutions will reestimate their models more frequently (monthly, weekly, or even daily). The daily reestimation using the moving window of the length of 1 000 observations is done in Kuester et al. (2006). Their results are much less positive than ours. As we have mentioned in section 2.1, they propose a new CAViaR specification and show that it performs much better than the other models.

Graphs of the 1%, resp. 5%, VaR estimates for S&P 500 are plotted in Figure 2.3, resp. 2.4 (SAV denotes the symmetric absolute model and AS denotes the asymmetric model). Graphs for the remaining indices can be found in the Appendix. Notice that the adaptive model produces the least volatile VaR forecasts. On the contrary, the indirect GARCH model's VaR estimates are the most volatile. Except for the adaptive model, all three remaining models produce extreme VaR estimates after the market crash in October 1987.

CAViaR models – 1% VaR – the first part.								
	Symmetric absolute value				Asymmetric slope			
	S&P	NASDAQ	DJ	NYSE	S&P	NASDAQ	DJ	NYSE
Beta1	0.133	0.341	0.199	0.178	0.188	0.167	0.383	0.393
Standard errors	0.035	0.034	0.045	0.020	0.042	0.039	0.118	0.151
p-values	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005
Beta2	0.921	0.811	0.879	0.893	0.855	0.811	0.716	0.690
Standard errors	0.016	0.011	0.027	0.022	0.025	0.040	0.089	0.134
p-values	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Beta3	0.184	0.466	0.300	0.252	*-0.029*	0.277	*0.083*	*0.080*
Standard errors	0.013	0.019	0.106	0.082	0.083	0.116	0.138	0.246
p-values	0.000	0.000	0.002	0.001	0.361	0.009	0.273	0.372
Beta4					0.522	0.577	0.763	*0.955*
Standard errors					0.107	0.170	0.283	0.793
p-values					0.000	0.000	0.003	0.114
RQ	193.223	204.980	182.378	175.643	184.994	197.853	175.035	169.191
Hits in-sample(%)	1.009	1.009	1.009	0.970	0.989	0.989	0.989	0.989
Hits out-of-sample(%)	0.600	0.600	0.700	0.600	0.500	0.800	0.900	0.700
DQ out-of-sample	0.007*	0.817	0.027	0.911	0.001*	0.714	0.056	0.979

Table 2.2: CAViaR models – 1% VaR – the first part. The first 5 054 observations were used to estimate the model and the remaining 1 000 observations were used for out-of-sample testing. Coefficients which are not significant at 5 % are *emphasized*; "*" denotes rejection according to the DQ test at 1% significance level.

CAViaR models – 1% VaR – the second part.								
	Indirect GARCH				Adaptive			
	S&P	NASDAQ	DJ	NYSE	S&P	NASDAQ	DJ	NYSE
Beta1	0.133	0.382	0.283	0.232	0.551	0.993	0.610	0.556
Standard errors	0.064	0.117	0.081	0.073	0.294	0.067	0.160	0.144
p-values	0.019	0.001	0.000	0.001	0.031	0.000	0.000	0.000
Beta2	0.923	0.798	0.847	0.848				
Standard errors	0.010	0.010	0.012	0.010				
p-values	0.000	0.000	0.000	0.000				
Beta3	*0.336*	*1.024*	*0.695*	*0.802*				
Standard errors	0.324	0.858	0.704	1.008				
p-values	0.150	0.116	0.162	0.213				
RQ	191.336	201.986	181.907	174.924	202.049	237.221	195.836	188.022
Hits in-sample(%)	1.049	1.029	1.029	1.009	0.970	0.989	0.970	0.989
Hits out-of-sample(%)	0.800	0.700	0.800	0.500	1.100	1.000	1.000	0.900
DQ out-of-sample	0.069	0.776	0.055	0.862	0.021	0.399	0.997	0.097

Table 2.3: CAViaR models – 1% VaR – the second part. The first 5 054 observations were used to estimate the model and the remaining 1 000 observations were used for out-of-sample testing. Coefficients which are not significant at 5 % are *emphasized*; "*" denotes rejection according to the DQ test at 1% significance level.

CAViaR models – 5% VaR – the first part.								
	Symmetric absolute value				Asymmetric slope			
	S&P	NASDAQ	DJ	NYSE	S&P	NASDAQ	DJ	NYSE
Beta1	0.034	0.070	0.068	0.049	0.027	0.054	0.077	0.064
Standard errors	0.004	0.007	0.016	0.006	0.009	0.015	0.019	0.016
p-values	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000
Beta2	0.958	0.924	0.927	0.941	0.936	0.859	0.871	0.879
Standard errors	0.008	0.008	0.022	0.012	0.013	0.024	0.032	0.023
p-values	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Beta3	0.089	0.156	0.117	0.109	*0.018*	0.115	*0.015*	*0.013*
Standard errors	0.018	0.018	0.034	0.025	0.019	0.035	0.022	0.031
p-values	0.000	0.000	0.000	0.000	0.160	0.001	0.252	0.332
Beta4					0.179	0.379	0.323	0.322
Standard errors					0.027	0.050	0.081	0.068
p-values					0.000	0.000	0.000	0.000
RQ	579.332	701.932	551.825	529.918	568.743	686.041	537.167	514.463
Hits in-sample(%)	5.046	5.046	5.065	5.046	5.046	5.026	5.006	5.046
Hits out-of-sample(%)	6.000	5.600	5.900	6.400	5.300	5.400	5.100	5.600
DQ out-of-sample	0.227	0.421	0.036	0.116	0.638	0.006*	0.150	0.111

Table 2.4: CAViaR models – 5% VaR – the first part. The first 5 054 observations were used to estimate the model and the remaining 1 000 observations were used for out-of-sample testing. Coefficients which are not significant at 5 % are *emphasized*; "*" denotes rejection according to the DQ test at 1% significance level.

CAViaR models – 5% VaR – the second part.								
	Indirect GARCH				Adaptive			
	S&P	NASDAQ	DJ	NYSE	S&P	NASDAQ	DJ	NYSE
Beta1	*0.020*	0.020	0.041	0.035	0.371	0.637	0.328	0.388
Standard errors	0.013	0.008	0.011	0.012	0.040	0.049	0.036	0.061
p-values	0.064	0.007	0.000	0.003	0.000	0.000	0.000	0.000
Beta2	0.937	0.909	0.920	0.913				
Standard errors	0.009	0.004	0.003	0.008				
p-values	0.000	0.000	0.000	0.000				
Beta3	*0.135*	*0.240*	*0.146*	*0.167*				
Standard errors	0.280	0.214	0.190	0.367				
p-values	0.315	0.131	0.222	0.324				
RQ	580.190	698.515	549.310	529.407	579.337	713.182	553.366	529.007
Hits in-sample(%)	5.125	5.026	5.065	5.065	4.749	4.769	4.511	4.610
Hits out-of-sample(%)	5.600	5.400	5.700	6.200	5.000	4.700	5.300	5.200
DQ out-of-sample	0.480	0.179	0.245	0.319	0.796	0.124	0.062	0.770

Table 2.5: CAViaR models – 5% VaR – the second part. The first 5 054 observations were used to estimate the model and the remaining 1 000 observations were used for out-of-sample testing. Coefficients which are not significant at 5 % are *emphasized*; "*" denotes rejection according to the DQ test at 1% significance level.

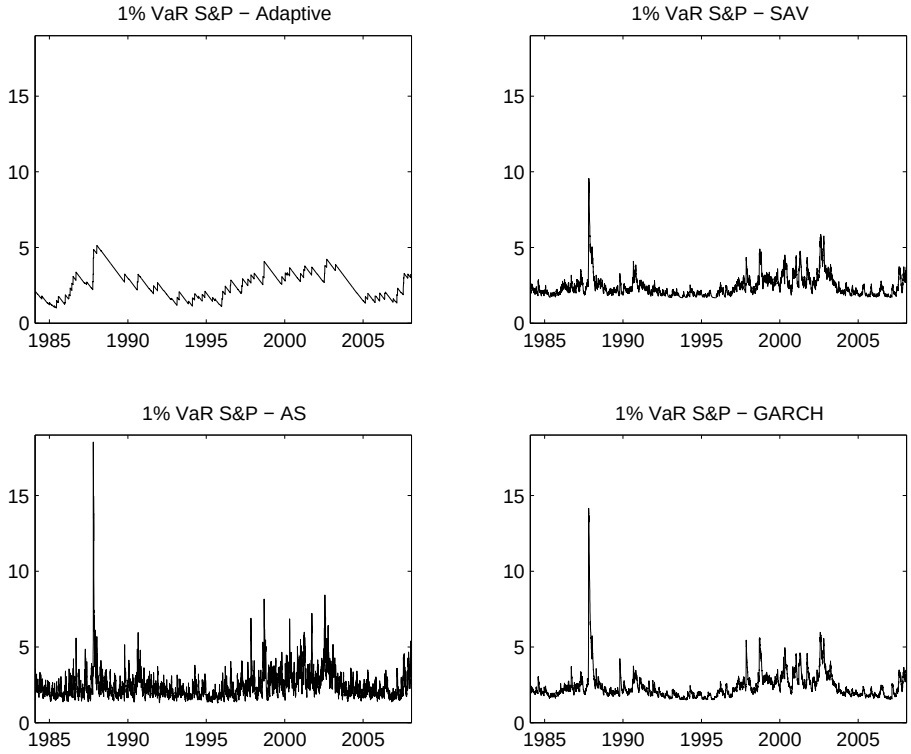


Figure 2.3: 1% VaR for S&P 500.

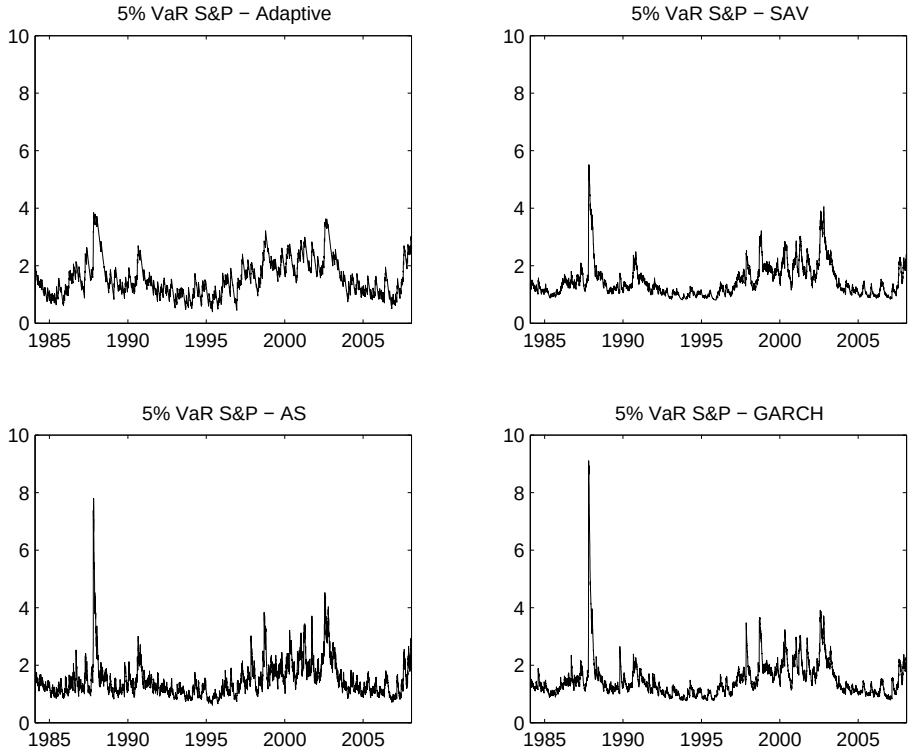


Figure 2.4: 5% VaR for S&P 500.

Chapter 3

Filtered historical simulation

3.1 Historical simulation

Probably the simplest method how to estimate VaR is the historical simulation (HS). Simplicity may be the reason why it is also one of the most popular approaches. As we have mentioned in the introductory chapter, the HS is nonparametric method, that means it makes no specific assumptions on the distribution of returns. The HS is based on the technique of rolling windows of observations of past returns. First, one need to choose the length of the window, which will be denoted as n. The one day θ VaR estimate for time $t+1$ is computed as the negative empirical θ-quantile estimator, $\widehat{Quant}_\theta(\cdot)$, of a moving window of n observations up to date t, that is, $\widehat{\text{VaR}}_{t+1} = -\widehat{Quant}_\theta(y_t, y_{t-1}, \ldots, y_{t-n+1})$. For instance, with a moving window of the length $n = 500$ observations, the 1% VaR estimate is simply the negative of the fifth sample order statistic.

This approach does not assume any specific distribution of returns, however it does assume that the returns are i.i.d., which implies that the distribution of returns does not change within the window. This assumption is often violated since financial returns exhibit nonconstant volatility. This may cause forecasting error. Imagine that the window is moving from an interval of low volatility to an interval of high volatility, then the VaR forecast based on the HS will underestimate the actual VaR. Conversely, if the window is moving from an interval of high volatility to an interval of low volatility, then the VaR forecast based on the HS will overestimate the actual VaR. This may be the reason why to choose shorter windows. However, extreme percentiles such as the 5% VaR and particularly the 1% VaR are very difficult to estimate accurately with small samples. In fact, the empirical quantile estimator is consistent only if the window size goes to infinity. Therefore, it is not easy to choose the appropriate length of the window since it must be large enough to estimate VaR accurately and it must not be too large in order to capture changes of volatility.

The choice of the length of the window is not the only doubtful part of this method. Another problematic characteristic is that VaR forecasts based on the HS may exhibit predictable jumps. When large negative returns enter into (resp. drop out

of) the window we can predict that VaR estimates will increase (resp. decrease). For instance, assume that we have a moving window of the length $n = 500$ and we estimate the 1% one day VaR. If the last five days exhibit extreme negative returns (compared with returns on the previous and subsequent 500 days) then it causes that the current VaR forecast increases and we also know that the VaR forecast will decrease after 500 days (when the extreme returns will be leaving the window). Last but not least, using the HS it is not possible to predict VaR that would exceed the extreme returns observed during the past n observations.

3.1.1 Empirical results

For the four indices described in section 2.4 we performed the historical simulation. We chose $\theta = 1\,\%$ and $5\,\%$ and estimated the one day VaR using the moving window of the length of 500, 1 000 and 1 500 observations. Therefore, we have 4 554 one day VaR predictions comprising the period from January 10, 1990, to February 1, 2008. Unlike the CAViaR method the computation of VaR estimates using the HS is very simple. All of the computations were done in MATLAB R2007b.

Table 3.1 presents results for all four indices, the percentage of hits and the corresponding p-values of the out-of-sample DQ test. Concerning the percentage of hits, we can notice that the VaR estimates underestimate the real VaR which results in higher percentage of hits than is expected. In the out-of-sample DQ test we used a constant, the VaR forecast and the first four lagged hits as the explanatory variables. For all four indices and all three window lengths the p-value of the out-of-sample DQ test is close to zero which means that the performance of the HS is unsatisfactory. The reason for rejection based on the DQ test may be the well known phenomenon of volatility clustering which means that large price changes are usually followed by more large changes. VaR is obviously affected by such an increase in volatility. Unfortunately, the HS method is not able to capture these short-term movements of volatility and produces violation clusters.

	1% VaR							
	S&P		NASDAQ		DJ		NYSE	
	Hits(%)	DQ	Hits(%)	DQ	Hits(%)	DQ	Hits(%)	DQ
500	1.340	*0.000*	1.098	*0*	1.361	*0.000*	1.318	*0.000*
1 000	1.296	*0.000*	1.361	*0*	1.230	*0.000*	1.296	*0.000*
1 500	1.186	*0.000*	1.493	*0*	1.120	*0.000*	1.318	*0.000*
	5% VaR							
	S&P		NASDAQ		DJ		NYSE	
	Hits(%)	DQ	Hits(%)	DQ	Hits(%)	DQ	Hits(%)	DQ
500	5.490	*0.000*	5.841	*0*	5.534	*0*	5.687	*0*
1 000	5.336	*0*	6.105	*0*	5.270	*0*	5.402	*0*
1 500	5.226	*0*	6.785	*0*	5.248	*0*	5.534	*0*

Table 3.1: Historical simulation – *emphasized* p-values denote rejection according to the DQ test at 1% significance level.

Graphs of one day VaR predictions for S&P 500 are plotted in Figure 3.1; solid lines correspond to the window of the length of 500 observations, dashed lines correspond to the window of the length of 1 000 observations and dotted lines correspond to the window of the length of 1 500 observations. Graphs of one day VaR predictions for S&P 500 together with S&P 500's losses are plotted in Figure 3.2. Analogous graphs for the remaining indices can be found in the Appendix. Notice that the shorter the window, the more volatile the VaR forecasts. This illustrates the fact that shorter windows capture changes of volatility better than longer windows. If we compare Figure 3.1 and Figure 2.2 we can see that from 1998 to 2004, when the returns of S&P 500 were more volatile, the VaR estimates are higher then for the rest of the period under examination (this is also evident in Figure 3.2 where, instead of returns, losses are plotted). After the year 2005 the VaR forecasts decrease. Notice that the decreasing comes first for the shortest window of 500 observations, then for the window of 1 000 observations and lastly for the longest window of 1 500 observations.[1] This is caused by the fact that if we have longer window it takes longer time for the large negative returns to leave the window. In Figure 3.2 we can see how often (and approximately when) losses exceeded VaR estimates.

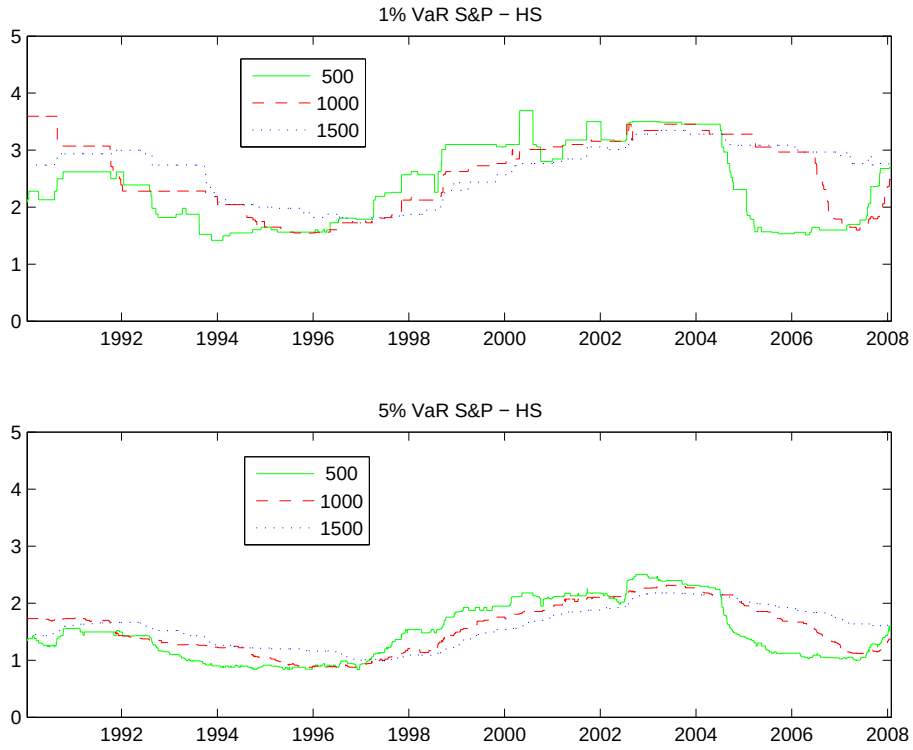


Figure 3.1: Historical simulation – VaR for S&P 500.

As we have mentioned above, the HS has some unsatisfactory properties. Hence, various authors have suggested different modifications of the HS method which should have better characteristics. Boudoukh et al. (1998) proposed a variant of

[1] This phenomenon can be seen especially well in Figure A.8 where the VaR predictions for NASDAQ are plotted.

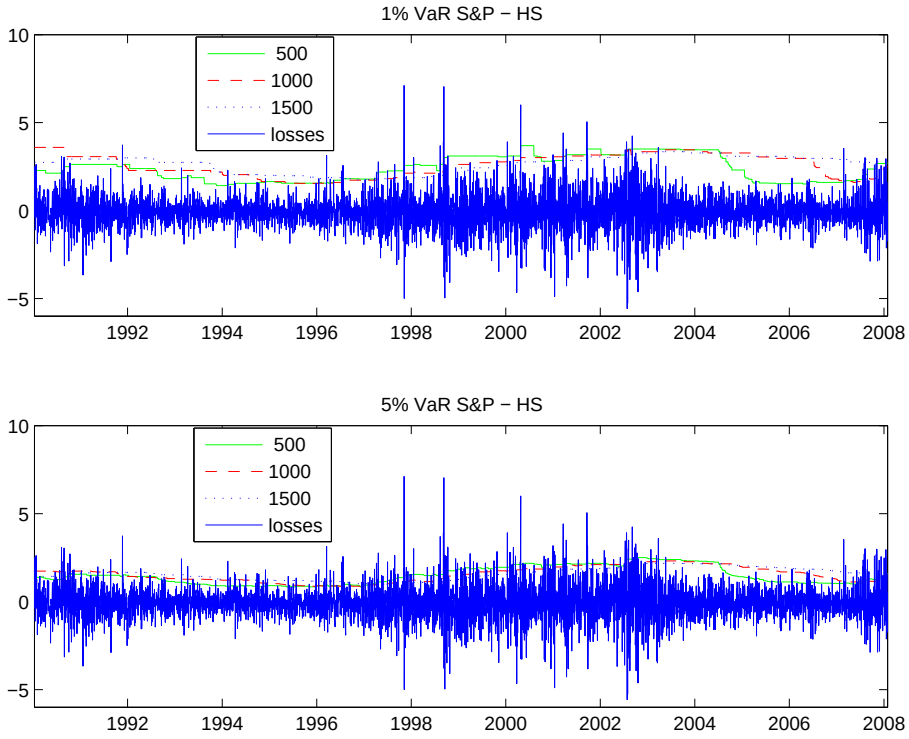


Figure 3.2: Historical simulation – VaR for S&P 500 compared with losses.

the HS which attempts to allow for stochastic volatility by sampling more frequently from recent observations than from observations generated in the distant past. This is done by applying exponentially declining probability weights (that sum to 1) to past returns. The weight given to the observation $k+1$ days ago is α times the weight given to the observation k days ago, where $0 < \alpha \leq 1$. VaR estimates are calculated based on the empirical cumulative function of returns with the modified probability weights. Note that when $\alpha = 1$ we get the HS method as a special case. For more details about this approach see Boudoukh et al. (1998), or Pritsker (2006) where some theoretical properties of this method are derived.

3.2 Filtered historical simulation

The HS method is based on the assumption that the returns are i.i.d., which allows to sample from past data with equal probabilities. However, this assumption is often violated. For instance, when volatilities change over time (which is typical for most daily financial time series) then the HS method can not adequately predict future losses. We will now describe the method of filtered historical simulation (FHS) developed by Barone-Adesi et al. (1999) which generalizes the HS and is designed to capture conditional heteroscedasticity.

In the FHS the stationarity assumption is relaxed. Barone-Adesi et al. (1999) propose to capture volatility clusters by modeling returns as GARCH processes (see Bollerslev (1986)). The simulation is based on the combination of GARCH

modeling (parametric approach) and historical returns series (non-parametric approach), therefore the whole procedure can be described as semi-parametric. The main idea is that historical residual returns are adapted to current market conditions by scaling them by the ratio of current over past conditional volatility.

We will describe the procedure on an example of an ARMA-GARCH(1,1) process, with the moving average term (ν) and the autoregressive term (μ)

$$y_t = \mu\ y_{t-1} + \nu\ \varepsilon_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, h_t), \tag{3.1}$$

$$h_t = \omega + \alpha\ \varepsilon_{t-1}^2 + \beta\ h_{t-1}, \tag{3.2}$$

where y_t is the return (at period t), ω, α and β are constants and ε_t is the random residual with the variance h_t. We form residual returns from the returns series by calibrating this GARCH model to the historical data. Historical residual returns $\hat{\varepsilon}_t$ are first standardized (divided) by volatility estimated on that particular day $\sqrt{\hat{h}_t}$. Thus, the standardized residual return e_t is given as

$$e_t = \frac{\hat{\varepsilon}_t}{\sqrt{\hat{h}_t}}.$$

This filtering process yields (under the GARCH hypothesis) i.i.d. residuals suitable for simulation. Standardized residuals e_t are scaled (multiplied) by the current conditional volatility forecast that reflects current market conditions. Then they are used as innovations in the conditional mean (3.1) and variance (3.2) equations to generate pathways for future returns and variances.

The standardized residual returns are drawn randomly from the data set and are afterwards used to form single pathways. The methodology of simulating single pathways proposed by Barone-Adesi et al. (1999) is the following. The first-drawn standardized residual e_1^* is scaled by the volatility forecast one day ahead to get the innovation forecast value for period $t+1$

$$z_{t+1} = e_1^*\ \cdot\ \sqrt{\hat{h}_{t+1}}, \tag{3.3}$$

where $\hat{h}_{t+1}$ is the variance forecast of period $t+1$ and can be calculated at the end of period t as

$$\hat{h}_{t+1} = \hat{\omega} + \hat{\alpha}\ \hat{\varepsilon}_t^2 + \hat{\beta}\ \hat{h}_t,$$

where $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ (as well as $\hat{\mu}$ and $\hat{\nu}$ in (3.4)) are the estimated parameters of the ARMA-GARCH model (that is the maximum likelihood estimates) and $\hat{\varepsilon}_t$ is the latest estimated residual return in (3.1). The innovation forecast z_{t+1} is then used to form the one-day ahead forecast of the asset returns

$$\hat{y}_{t+1} = \hat{\mu}\ y_t + \hat{\nu}\ \hat{\varepsilon}_t + z_{t+1}. \tag{3.4}$$

For $i = 2, 3, \ldots$ (to the length of the holding period), the volatility for period $t+i$ is unknown (at the end of period t) and must be simulated. Forecasts of volatility are simulated by the recursive substitution of scaled residuals into the variance

equation (3.2). Thus the first-drawn standardized residual with which we form the return forecast one-day ahead in (3.4) also allows for the simulation of the volatility forecast two days ahead

$$\sqrt{\hat{h}_{t+2}} = \sqrt{\hat{\omega} + \hat{\alpha}\ z_{t+1}^2 + \hat{\beta}\ \hat{h}_{t+1}}.$$

Then the second-drawn standardized residual e_2^* is scaled by this volatility forecast $\sqrt{\hat{h}_{t+2}}$ to get the innovation forecast value for period $t+2$, that is

$$z_{t+2} = e_2^*\ \cdot\ \sqrt{\hat{h}_{t+2}},$$

and is used to simulate the return two days ahead

$$\hat{y}_{t+2} = \hat{\mu}\ \hat{y}_{t+1} + \hat{\nu}\ z_{t+1} + z_{t+2}.$$

Analogously the volatility forecast three days ahead is formed from the previous second-drawn scaled residual and allows for the scaling of the third-drawn residual e_3^* and so on up to the length of the holding period.

In general, the simulated volatility forecast for period $t+i$ ($\sqrt{\hat{h}_{t+i}}$) is obtained as

$$\sqrt{\hat{h}_{t+i}} = \sqrt{\hat{\omega} + \hat{\alpha}\ z_{t+i-1}^2 + \hat{\beta}\ \hat{h}_{t+i-1}}, \qquad i \geq 2,$$

where the innovation forecasts z_{t+k} are estimated similarly as in (3.3), that is

$$z_{t+k} = e_k^*\ \cdot\ \sqrt{\hat{h}_{t+k}}, \qquad k \geq 1,$$

where e_k^* is the k^{th}-drawn standardized residual. The simulated returns $\hat{y}_{t+i}$ are given as

$$\hat{y}_{t+i} = \hat{\mu}\ \hat{y}_{t+i-1} + \hat{\nu}\ z_{t+i-1} + z_{t+i}, \qquad i \geq 2.$$

By repeating this procedure a large number of times we form many pathways and obtain the "empirical" distribution of simulated returns at the chosen holding period for a single asset. VaR forecast is then computed as the negative quantile of this distribution. The simulation methodology can be extended for a portfolio of multiple assets (derivative securities) as is done in Barone-Adesi et al. (1999).

ARMA-GARCH(1,1) process described above is not the only possible GARCH specification which can be used. One can use various other GARCH specifications or even other time series models that are appropriate for the data under examination. Also the assumption that the residual returns follow the normal distribution can be modified by replacing the normal distribution with different distributions to achieve better filtering, for example with the Student's t-distribution (see Kuester et al. (2006)). The only requirement is that the model generates i.i.d. residuals from the return series and that all parameters in the model can be consistently estimated.

The great advantage of the FHS over the HS is that the filtering process increases the range of outcomes beyond the historical record through a change of scale. This means that the FHS provides an approach to generate extreme events which are not present in the historical record, completing the tails of the distribution. Beside Barone-Adesi et al. (1999), also Barone-Adesi et al. (2002), Pritsker (2006) and Kuester et al. (2006) show that the FHS method gives good results and outperforms not only the HS, but other methods as well.

3.3 Volatility updating

We will discuss the method of incorporating volatility updating into the HS proposed by Hull and White (1998) which improves the HS by taking account of short-term movements of volatility. This method is essentially identical to the FHS method when the one day VaR is estimated (i.e. the holding period is equal to one day). The method is based on the assumption that when the returns are scaled by an estimate of volatility then the probability distribution of scaled returns is stationary.

The basic idea is analogous to the FHS, that is the historical returns are adapted to current market conditions by scaling them by the ratio of current over past conditional volatility. Let y_t be the return at time t and $\hat{\sigma}_t$ be the estimate of the daily volatility of the return for time t made at the end of day $t-1$. Then it is assumed that the probability distribution of scaled returns $y_t/\hat{\sigma}_t$ is stationary. Consider that we are estimating (at time s) the one day VaR for time $s+1$. Then the most recent estimator of the daily volatility is $\hat{\sigma}_{s+1}$, this is the estimate of the volatility of the return during day $s+1$ made at the end of day s. We replace every historical return y_t by the return with updated volatility y_t^{*s} which is computed as follows

$$y_t^{*s} = \hat{\sigma}_{s+1} \frac{y_t}{\hat{\sigma}_t}, \quad t \leq s. \tag{3.5}$$

If we chose n to be the length of the window, then the θ VaR estimate for time $s+1$ is computed as the negative empirical θ-quantile estimator, $\widehat{Quant}_\theta(\cdot)$, of a moving window of n returns with updated volatility y_t^{*s} up to date s, that is

$$\widehat{\mathrm{VaR}}_{s+1} = -\widehat{Quant}_\theta(y_s^{*s}, y_{s-1}^{*s}, \ldots, y_{s-n+1}^{*s}). \tag{3.6}$$

That means, instead of using the actual historical returns for computing VaR predictions we use returns with updated volatility that reflect the ratio of the current daily volatility to the daily volatility at the time of the particular observation. To compute the VaR prediction the following day, the whole window is moved forward by one observation and the procedure is repeated.

The appealing feature of incorporating volatility updating into the HS method is that it allows for time-varying volatility but the distribution of the returns is modeled nonparametrically. Moreover, it also allows for VaR estimates that exceed the historical extreme observations of returns. (On the contrary, the HS method is unable to predict losses which are outside the sample of historical observations.)

Estimation of volatility

Different methods can be used for estimating the daily volatility, for instance, one can compute the volatility estimator as the square root of the sample variance over the estimation window or use a GARCH model. We will use the method proposed in RiskMetrics (see J.P. Morgan (1995)).

The RiskMetrics method estimates the returns' volatility using an exponentially weighted moving average (EWMA) model for the daily variances, that is

$$\hat{\sigma}_t^2 = \sum_{i=1}^{\infty} \omega_i (y_{t-i} - \bar{y})^2, \tag{3.7}$$

where $\hat{\sigma}_t^2$ is the estimated variance calculated at time $t-1$, y_t is the return (at time t), $\bar{y}$ is the estimated mean return, the weights ω_i are given as

$$\omega_i = (1-\lambda)\lambda^{i-1} \tag{3.8}$$

and parameter λ, $0 < \lambda < 1$, is an decay factor. The EWMA model attach different weights to the past observations. Since the weights decline exponentially, the most recent observations receive much more weight than earlier observations.

The RiskMetrics model uses $\lambda = 0.94$ for daily financial data sets and assumes that the mean return $\bar{y}$ is negligible and sets it to zero.[2] If we insert (3.8) into (3.7) and set $\bar{y} = 0$, we get

$$\begin{aligned} \hat{\sigma}_t^2 &= \sum_{i=1}^{\infty}(1-\lambda)\lambda^{i-1} y_{t-i}^2 = \sum_{i=2}^{\infty}(1-\lambda)\lambda^{i-1} y_{t-i}^2 + (1-\lambda) y_{t-1}^2 \\ &= \lambda \sum_{i=1}^{\infty}(1-\lambda)\lambda^{i-1} y_{t-1-i}^2 + (1-\lambda) y_{t-1}^2 = \lambda \hat{\sigma}_{t-1}^2 + (1-\lambda) y_{t-1}^2. \end{aligned}$$

and the estimation formula of EWMA model (3.7) can be rewritten as an easy recursive formula

$$\hat{\sigma}_t^2 = \lambda \hat{\sigma}_{t-1}^2 + (1-\lambda) y_{t-1}^2.$$

The estimate of variance on any given day is a combination of two components: the estimate of variance on the previous day, which receives the weight λ, and the squared return on the previous day, which receives the weight $(1-\lambda)$.

The estimator of volatility for time t under the EWMA model is computed as

$$\hat{\sigma}_t = \sqrt{\lambda \hat{\sigma}_{t-1}^2 + (1-\lambda) y_{t-1}^2}. \tag{3.9}$$

The starting value of volatility $\hat{\sigma}_1$ and λ are parameters of the model. In the following subsection 3.3.1 we set $\lambda = 0.94$ and $\hat{\sigma}_1 = 1$, and computed the volatility estimates using the equation (3.9). Note that the computations were done

[2]The mean of the returns is assumed to be zero in many cases. For comparison of results when mean adjustment is or is not considered see Hull and White (1998).

in MATLAB R2007b and both of the parameters can easily be changed at the beginning of the programs.

The computation of VaR estimates using the HS with volatility updating is very simple (almost as simple as the HS method itself). To summarize briefly the volatility updating procedure, we will now describe how the computations proceed:

1. Consider a total of T observations of historical daily returns $y_1, \ldots, y_T$. (We have $T = 6\,054$.)
2. The length of the window n and the probability θ is chosen. (We chose n to be 500, 1 000, 1 500 and θ to be 1 % and 5 %.)
3. Starting value $\hat{\sigma}_1$ and λ are set. (We set $\lambda = 0.94$ and $\hat{\sigma}_1 = 1$.)
4. Set $s = n$.
5. Consider the window of observed historical returns $y_{s-n+1}, \ldots, y_s$.
6. For $t = s - n + 1, \ldots, s$ update $\hat{\sigma}_t$ according to (3.9) and compute returns with updated volatility y_t^{*s} according to (3.5).
7. Compute $\widehat{\mathrm{VaR}}_{s+1}$, that is the estimate of the θ VaR at time $s+1$ (estimated at time s), according to (3.6).
8. Unless $s = T - 1$, set $s = s + 1$ and move to the item 5. (Shift of the window provides a newly included return and allows for updating of volatility and returns according to the item 6.)

3.3.1 Empirical results

For the four indices described in section 2.4 we performed the historical simulation with volatility updating. We again chose $\theta = 1\,\%$ and $5\,\%$ and estimated the one day VaR using the moving windows of the same lengths as in the HS method in section 3.1 (that is 500, 1 000 and 1 500 observations). Therefore, for all three window lengths we have 4 554 one day VaR predictions comprising the period from January 10, 1990, to February 1, 2008 which we can compare.

Table 3.2 presents results for all four indices, the percentage of hits and the corresponding p-values of the out-of-sample DQ test (in the out-of-sample DQ test we used a constant, the VaR forecast and the first four lagged hits as the explanatory variables). We can compare these results with results of the HS method which are presented in Table 3.1. First, we can notice that this time the percentage of hits is closer (than for the HS) to the expected 1 %, resp. 5 %, that means we get better results than for the HS. Recall that the performance of the HS was unsatisfactory with respect to the out-of-sample DQ test. Using volatility updating we achieved an improvement, for all four indices at least one window length is not rejected by the out-of-sample DQ test in the case of 1% VaR (for NYSE even all window lengths are not rejected). In the case of 5% VaR the p-value of the out-of-sample

DQ test is close to zero for all four indices and all three window lengths, only for the S&P 500 the window of the length of 1 500 observations is not rejected. To summarize, the HS with volatility updating obviously outperforms the HS method regarding the percentage of hits and the out-of-sample DQ test.

	1% VaR							
	S&P		NASDAQ		DJ		NYSE	
	Hits(%)	DQ	Hits(%)	DQ	Hits(%)	DQ	Hits(%)	DQ
500	0.922	0.022	0.988	0.165	0.922	*0.009*	0.922	0.042
1 000	1.120	*0.001*	1.010	*0.000*	1.032	0.017	0.988	0.013
1 500	1.120	*0.001*	0.988	*0.000*	1.098	*0.005*	0.966	0.011
	5% VaR							
	S&P		NASDAQ		DJ		NYSE	
	Hits(%)	DQ	Hits(%)	DQ	Hits(%)	DQ	Hits(%)	DQ
500	5.314	*0.000*	4.941	*0.000*	5.051	*0.000*	5.270	*0.000*
1 000	5.094	*0.005*	4.963	*0.001*	5.182	*0.000*	5.270	*0.000*
1 500	5.094	0.012	4.985	*0.008*	5.160	*0.000*	5.226	*0.000*

Table 3.2: The HS with updated volatility – *emphasized* p-values denote rejection according to the DQ test at 1% significance level.

Graphs of one day VaR predictions for S&P 500 are plotted in Figure 3.3; solid lines correspond to the window of the length of 500 observations, dashed lines correspond to the window of the length of 1 000 observations and dotted lines correspond to the window of the length of 1 500 observations. Graphs of one day VaR predictions for S&P 500 together with S&P 500's losses are plotted in Figure 3.4. Analogous graphs for the remaining indices can be found in the Appendix. The first thing we can notice in Figure 3.3 is that (unlike the HS method) there is very small difference between VaR forecasts using different window lengths.[3] If we compare Figure 3.3 and Figure 3.1 we can see that when the HS with volatility updating is used then the VaR estimates are more volatile. This is not surprising since the EWMA model clearly aims to capture short-term movements of volatility. In figure Figure 3.4 we can observe how the VaR estimates change according to the changes of losses and how often (and approximately when) losses exceeded VaR estimates. Notice that some VaR predictions exceed the previous extreme observations of returns. We can also compare Figure 3.4 and Figure 3.2. In Figure 3.2 we observe that the HS method is not able to capture short-term movements of volatility quickly enough. On the other hand, Figure 3.4 shows that the HS with volatility updating immediately responds to changes of volatility.

[3]The difference can be seen better in colored Figures in the DiplomaThesisC.ps file which is on the enclosed CD.

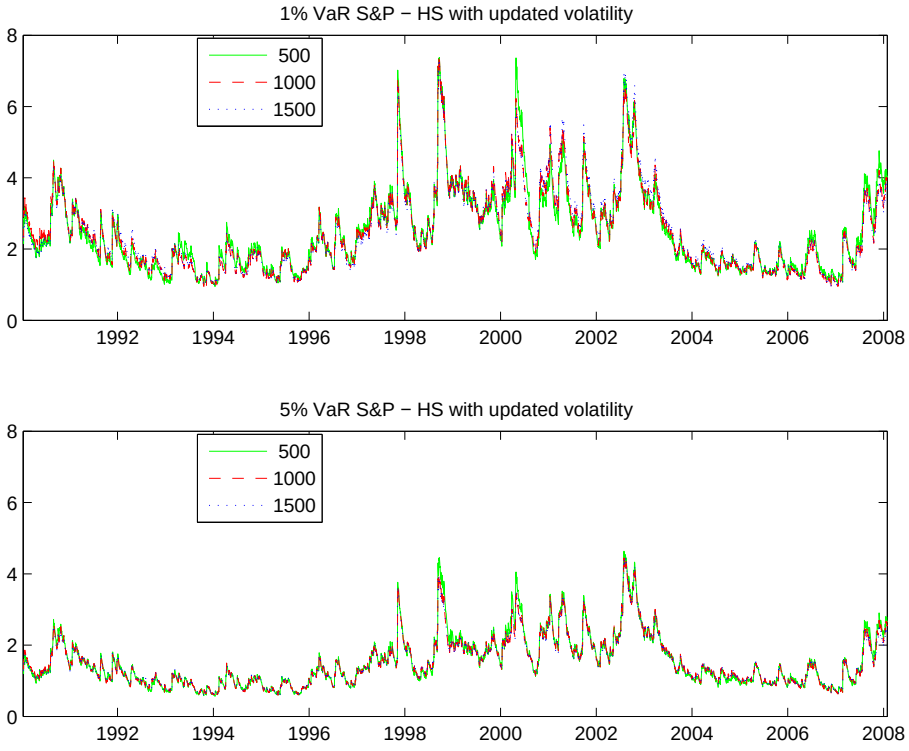


Figure 3.3: The HS with updated volatility – VaR for S&P 500.

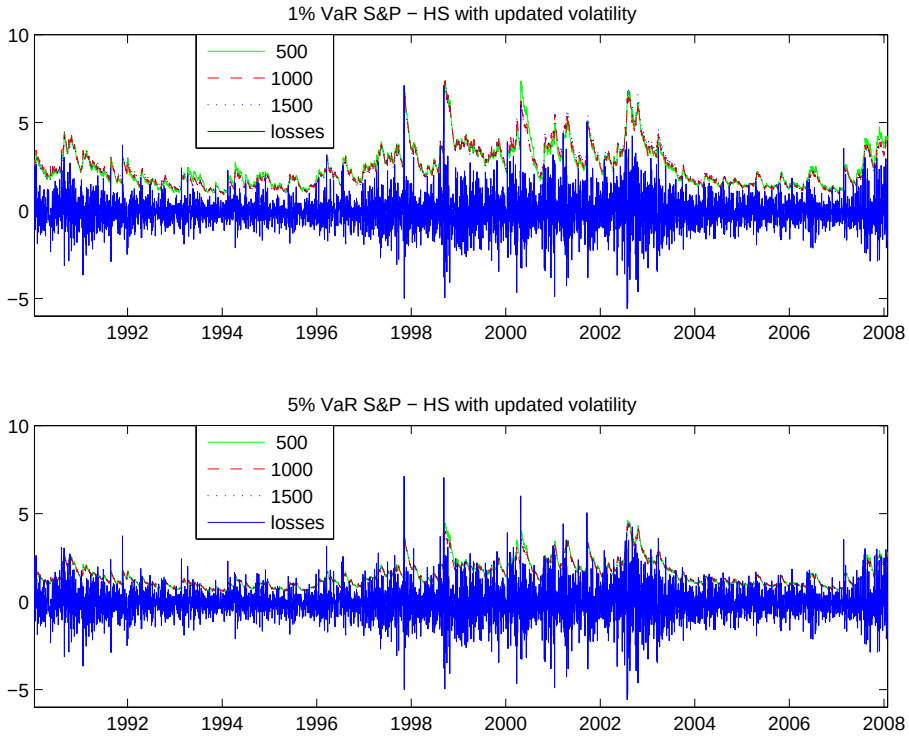


Figure 3.4: The HS with updated volatility – VaR for S&P 500 and observed losses.

Chapter 4

Conclusion

One of the most important tasks of financial institutions is evaluation of exposure to the market risk which arises from variations in prices. A commonly used methodology for estimation of the market risk is the VaR concept. The traditional approaches to VaR computations based on assumptions of i.i.d.-ness and normality do not provide satisfactory quantification of possible losses since the financial data usually posses fatter tails. Therefore, many alternative approaches have been suggested. We discussed two alternative approaches in more details.

First, we described the CAViaR method together with its asymptotic properties. The empirical experiments with real data showed that this method is able to give good results considering the percentage of hits (i.e. the percentage of VaR exceeding) and the out-of-sample DQ test. The main disadvantage of this method seems to be computational demands.

On the contrary, one of the reasons why historical simulation based methods for computing VaR became popular is that these methods are simple and easy to implement. The empirical results that we obtained using the HS method were unsatisfactory regarding the percentage of hits and the out-of-sample DQ test. The HS method with updated volatility (which is a special case of the filtered historical simulation for one day VaR prediction) is an easy extension of the HS method and gave better results considering the percentage of hits and the out-of-sample DQ test.

Even though VaR predictions of the FHS method and the CAViaR method seem to be more accurate than VaR predictions of the HS method, their greater volatility may be undesirable. For instance, when VaR is used for calculation of regulatory market-risk capital requirements, then the financial institution would probably rather keep stable amount of the capital which should cover potential losses. As a consequence, deciding which method for estimating VaR should be used is not an easy task and the choice is influenced by many aspects (not only by the accuracy of VaR predictions, but also by the purpose of VaR estimation, by the character of available data, etc.).

Appendix A

Graphs and tables

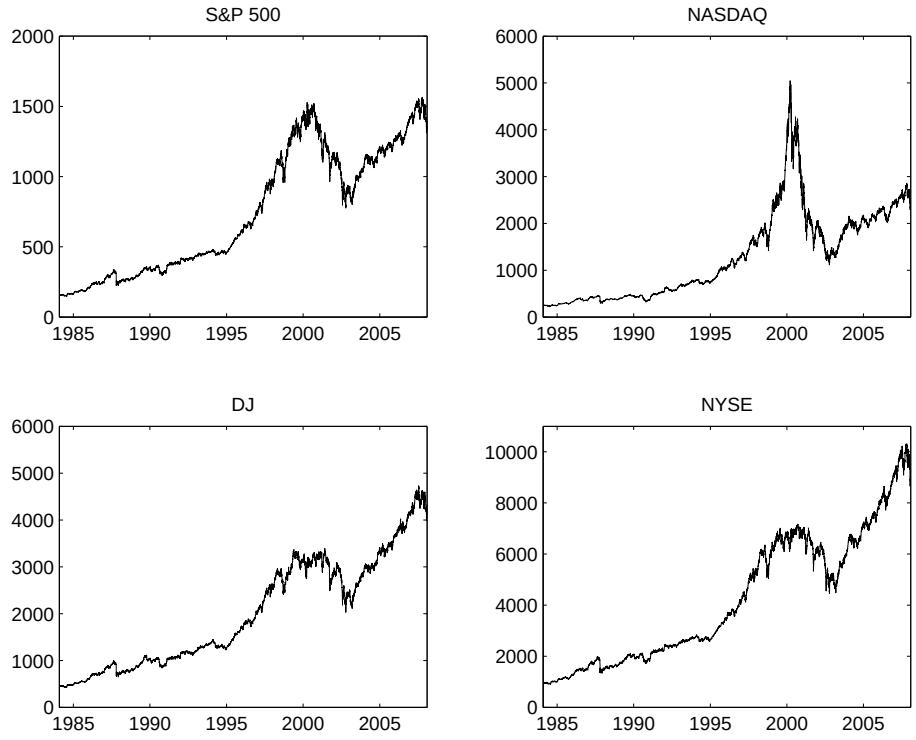


Figure A.1: Daily prices of indices – graphs have different scales on the y-axis.

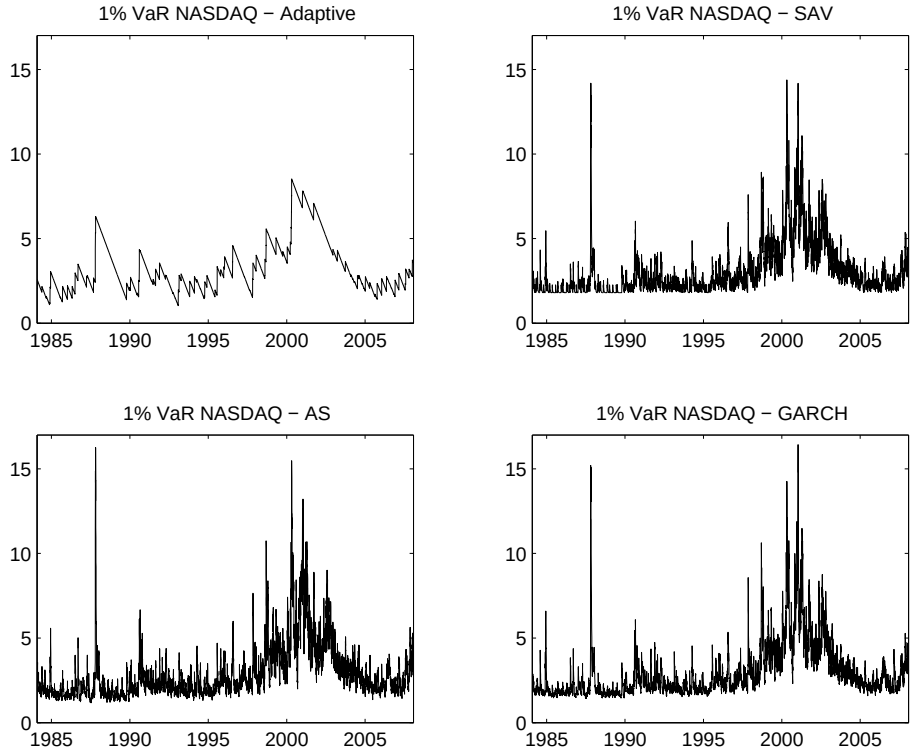


Figure A.2: 1% VaR for NASDAQ.

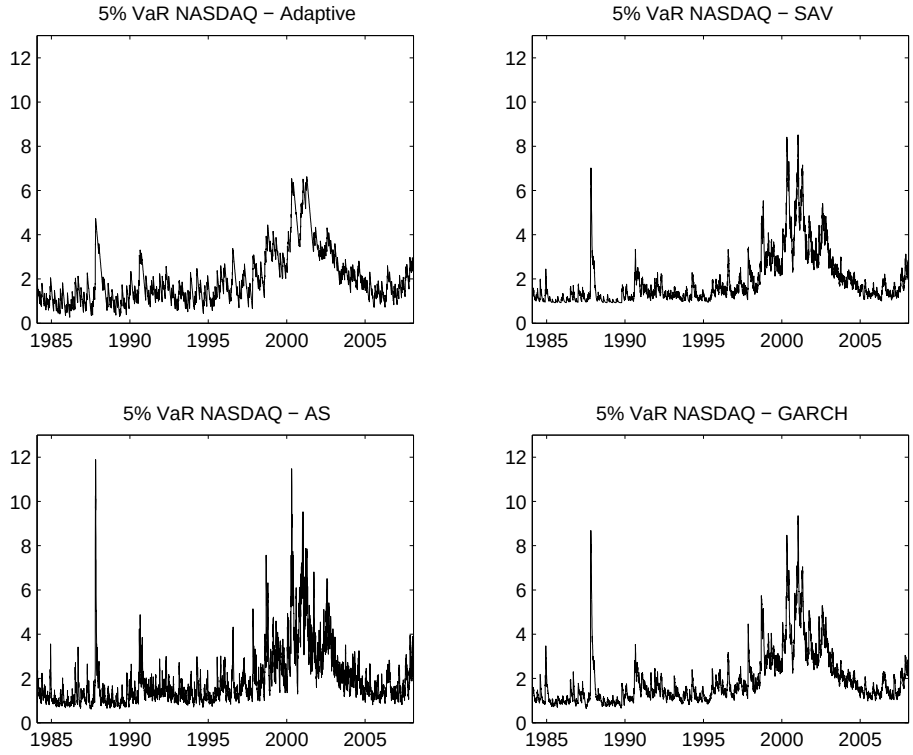


Figure A.3: 5% VaR for NASDAQ.

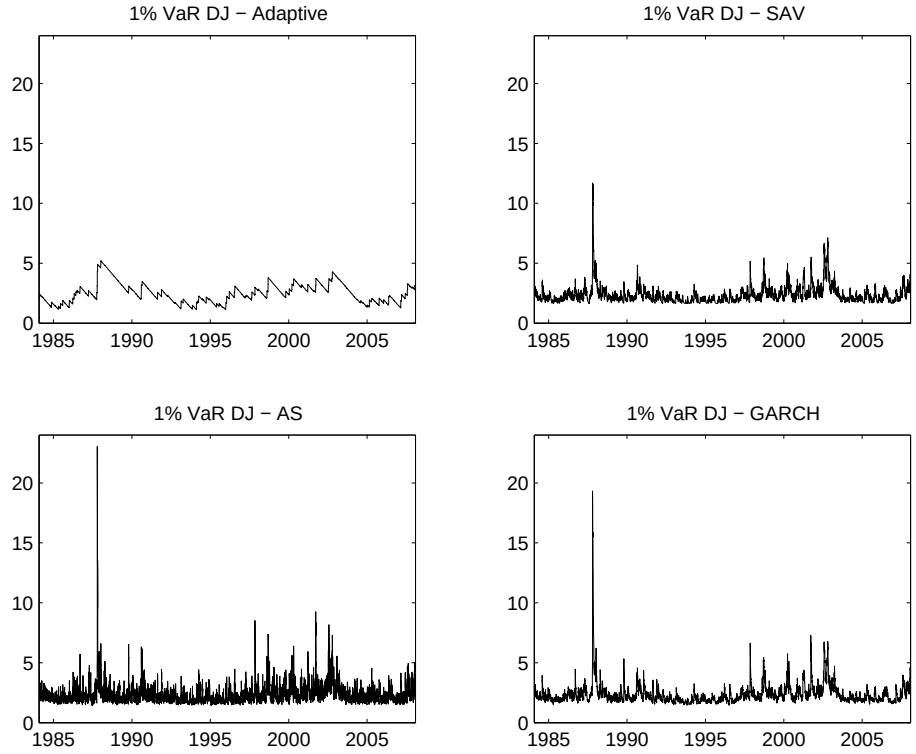


Figure A.4: 1% VaR for Dow Jones Composite index.

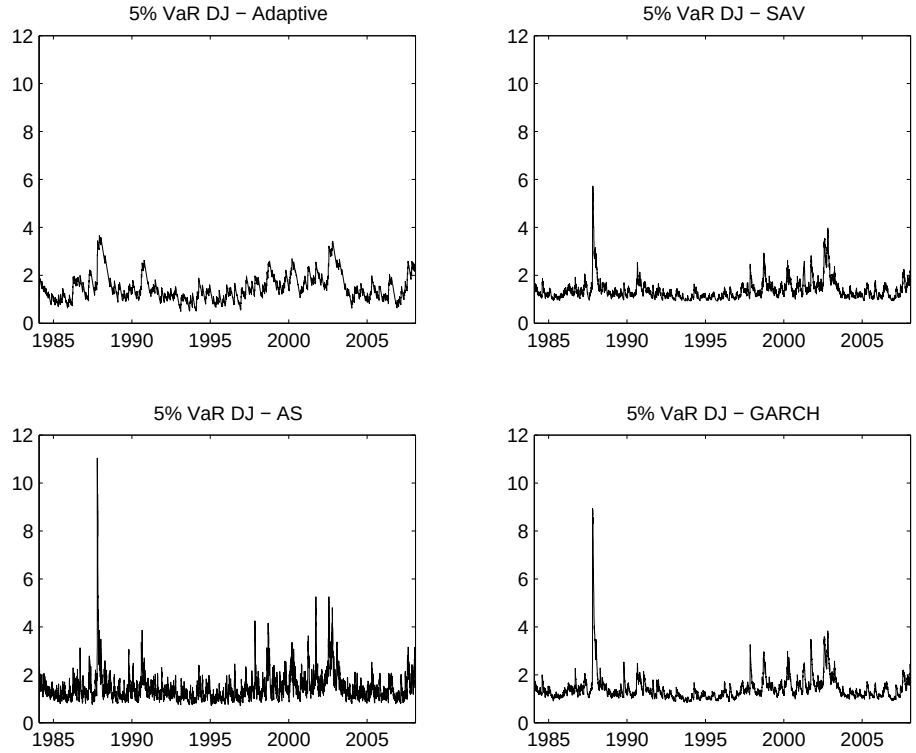


Figure A.5: 5% VaR for Dow Jones Composite index.

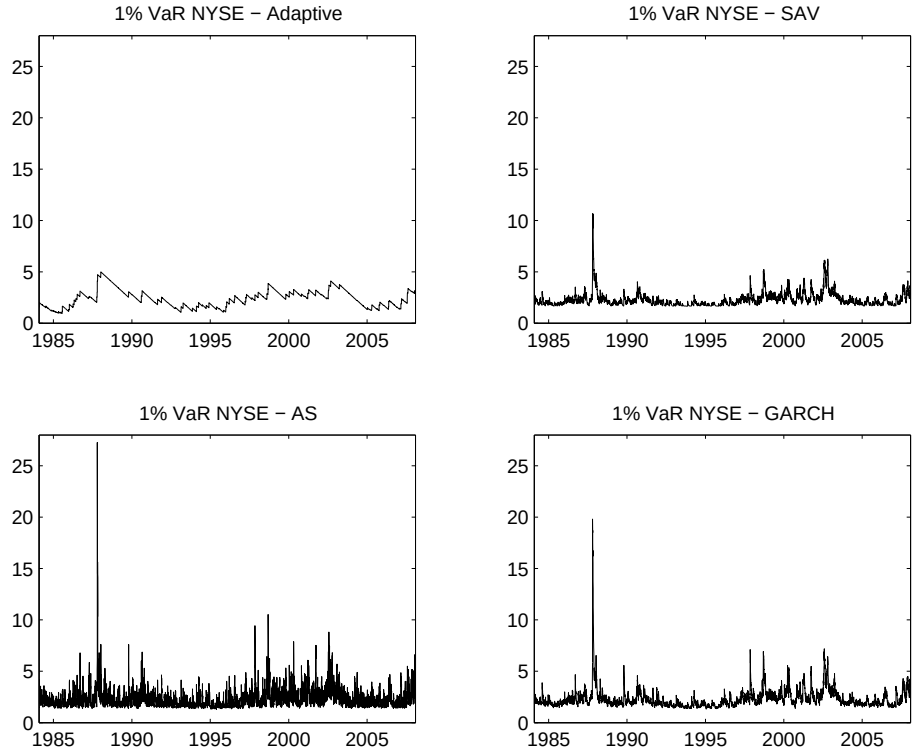


Figure A.6: 1% VaR for NYSE.

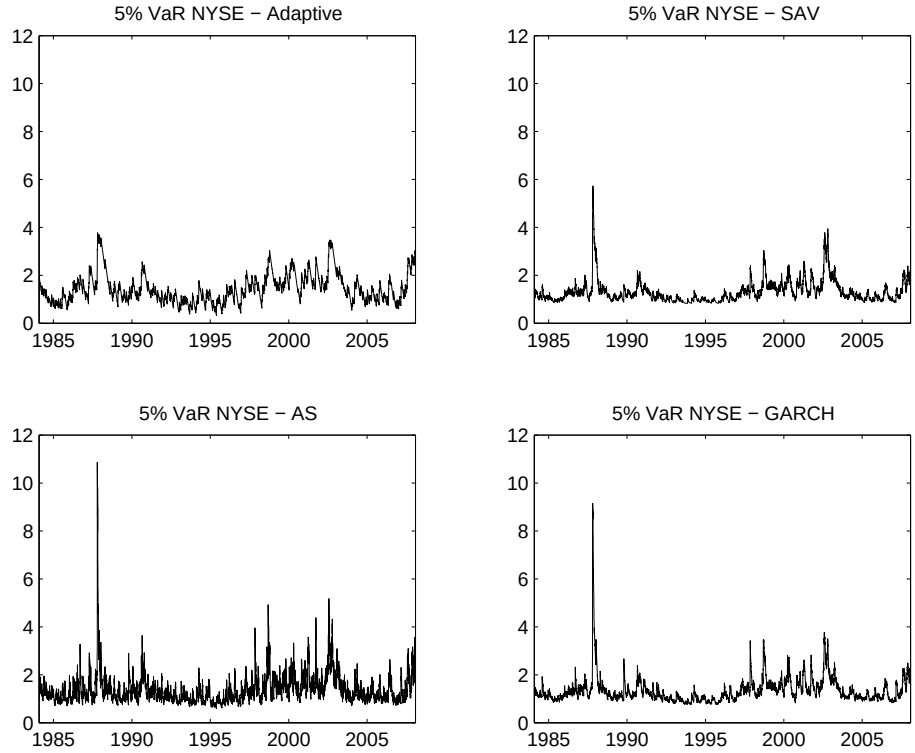


Figure A.7: 5% VaR for NYSE.

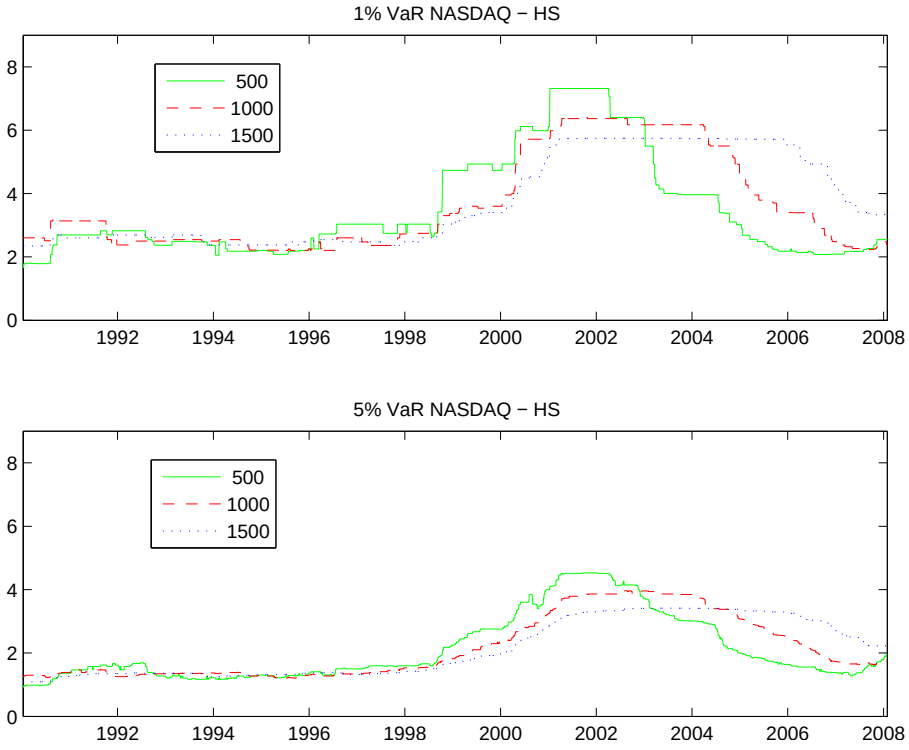


Figure A.8: Historical simulation – VaR for NASDAQ.

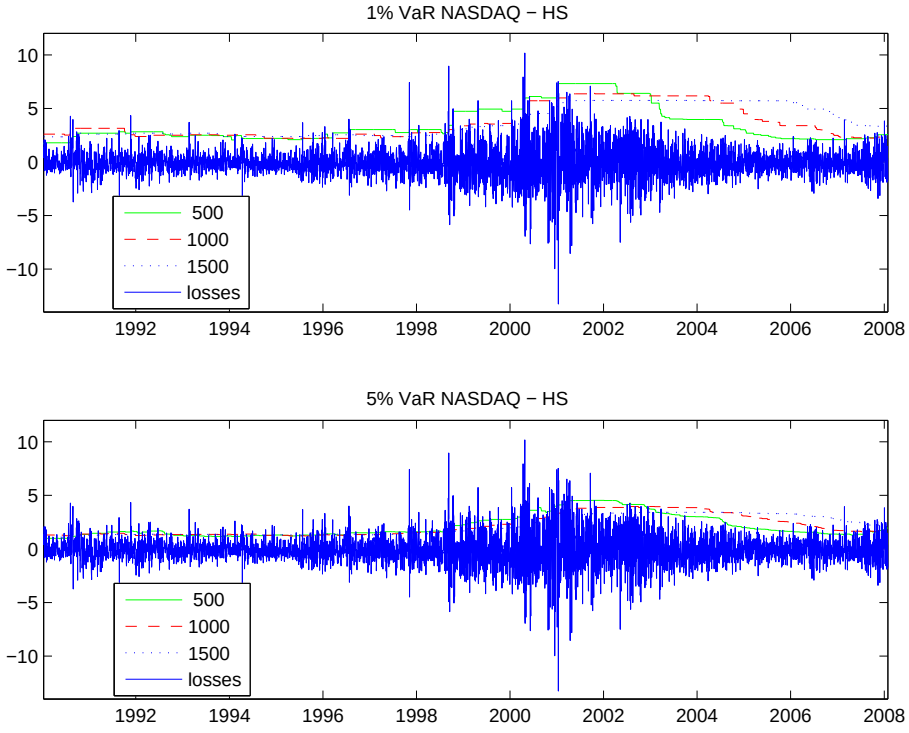


Figure A.9: Historical simulation – VaR for NASDAQ compared with losses.

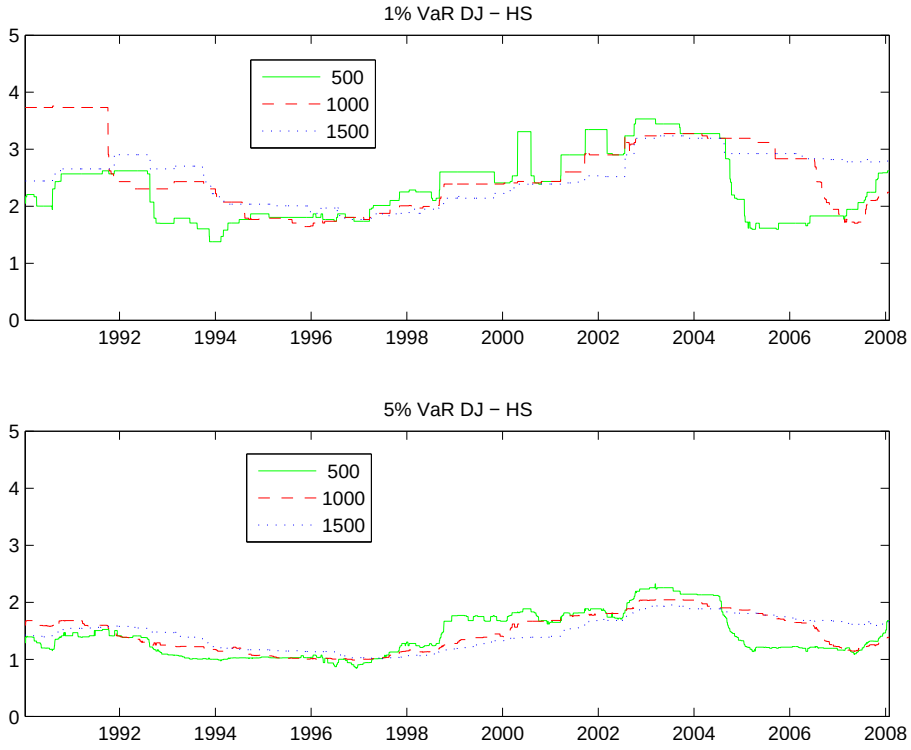


Figure A.10: Historical simulation – VaR for Dow Jones.

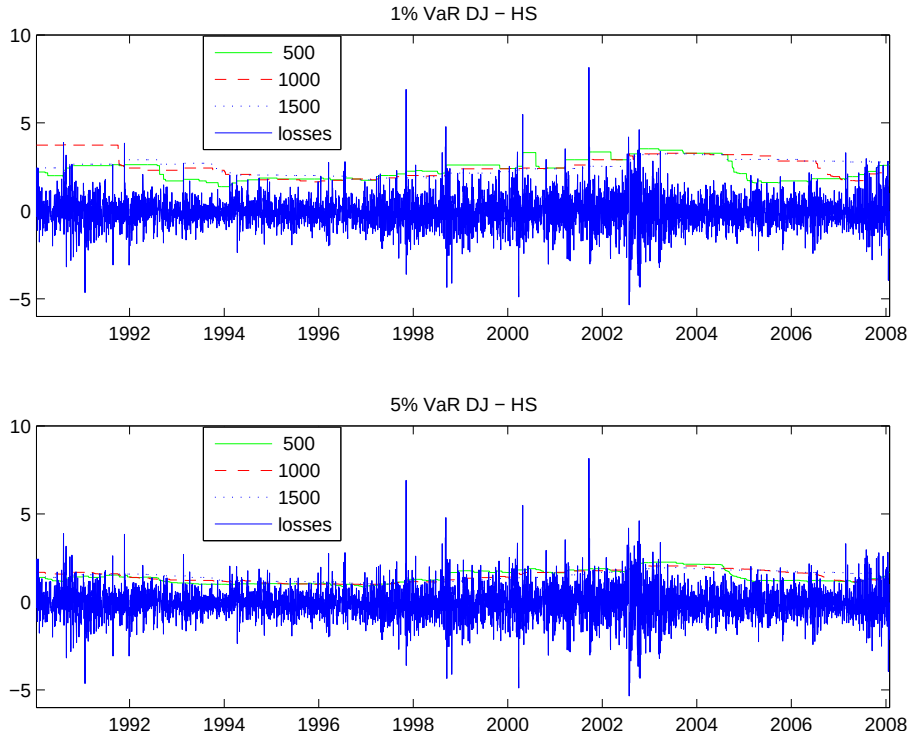


Figure A.11: Historical simulation – VaR for Dow Jones compared with losses.

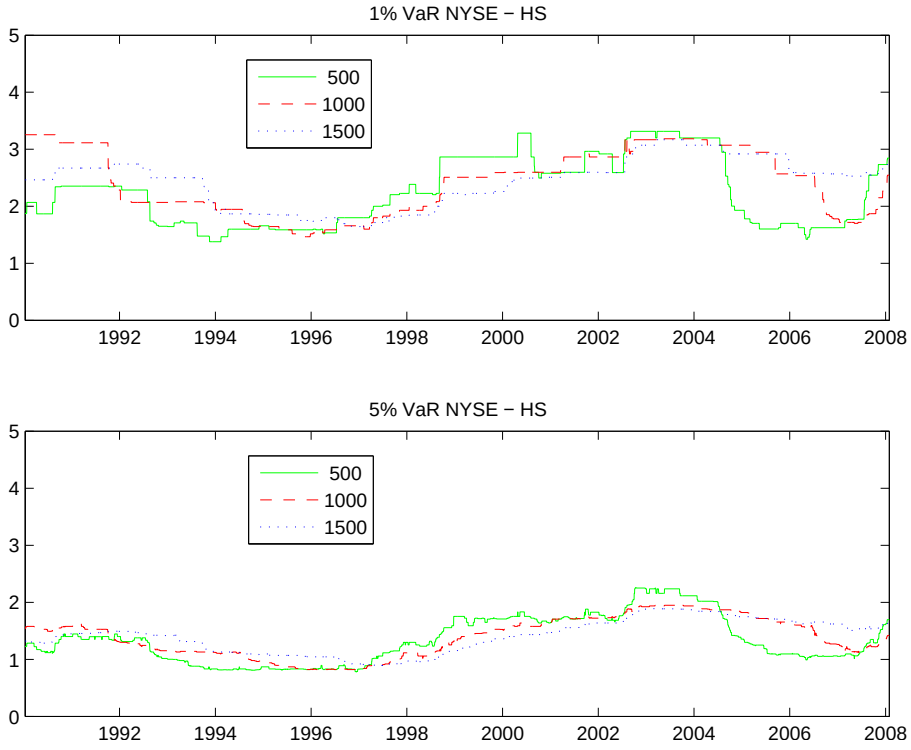


Figure A.12: Historical simulation – VaR for NYSE.

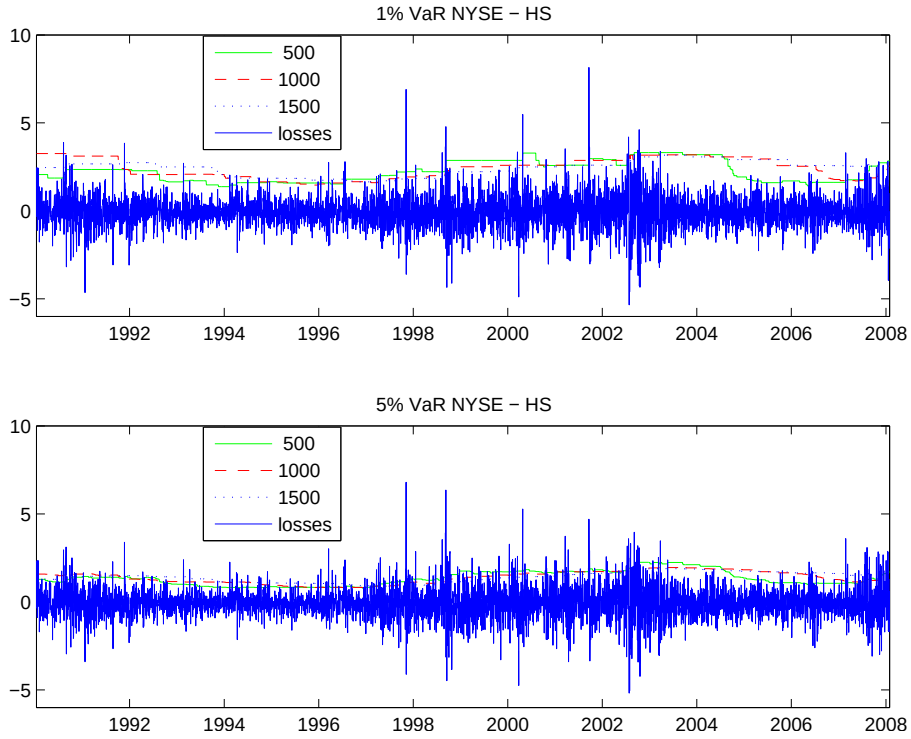


Figure A.13: Historical simulation – VaR for NYSE compared with losses.

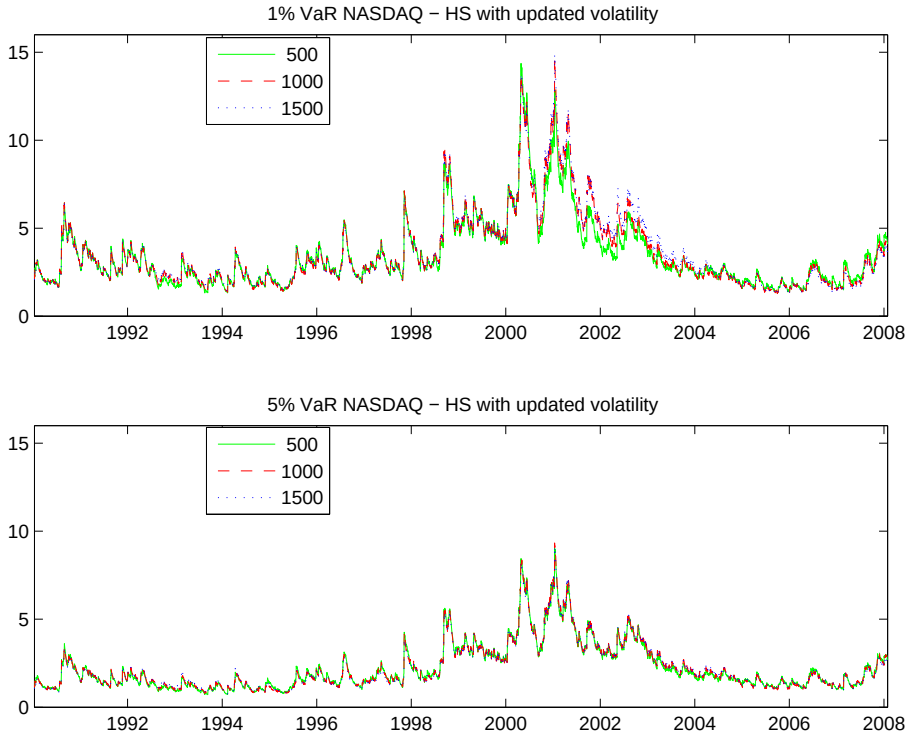


Figure A.14: The HS with updated volatility – VaR for NASDAQ.

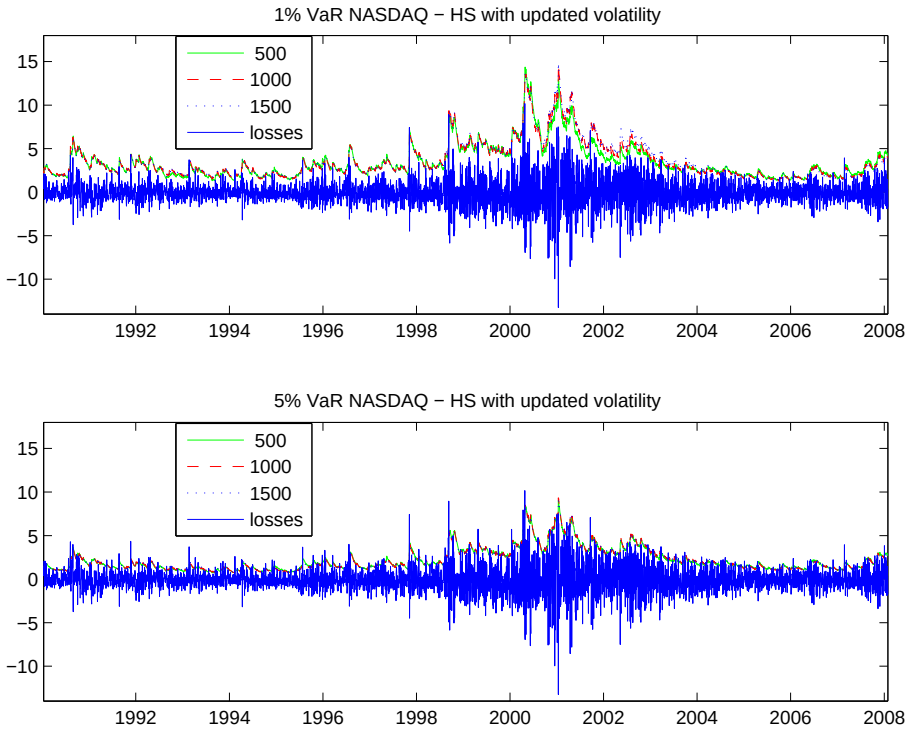


Figure A.15: The HS with updated volatility – VaR for NASDAQ and losses.

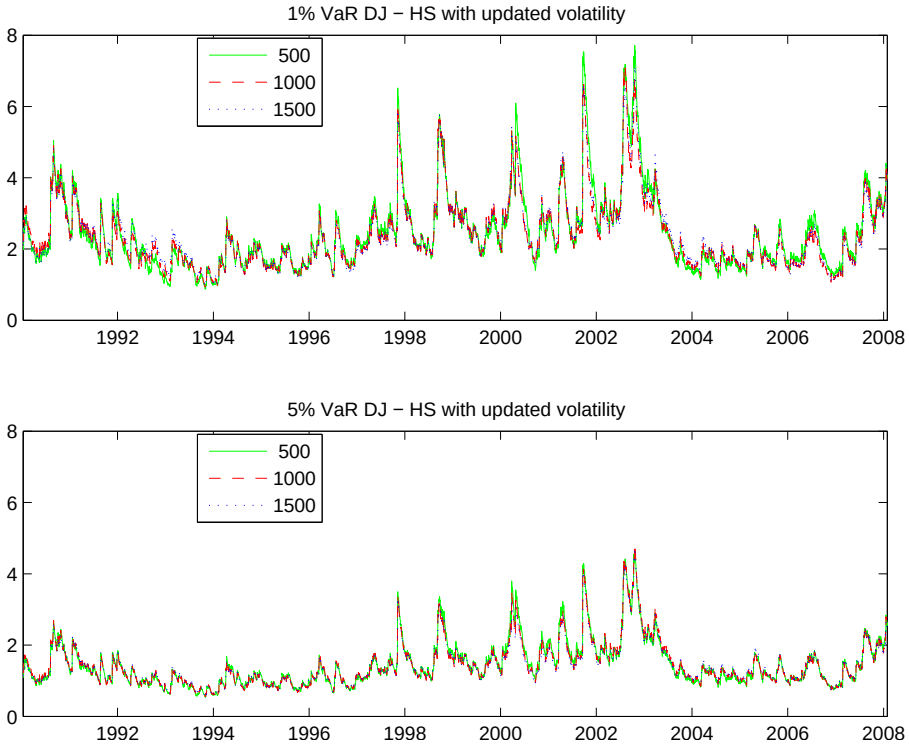


Figure A.16: The HS with updated volatility – VaR for Dow Jones.

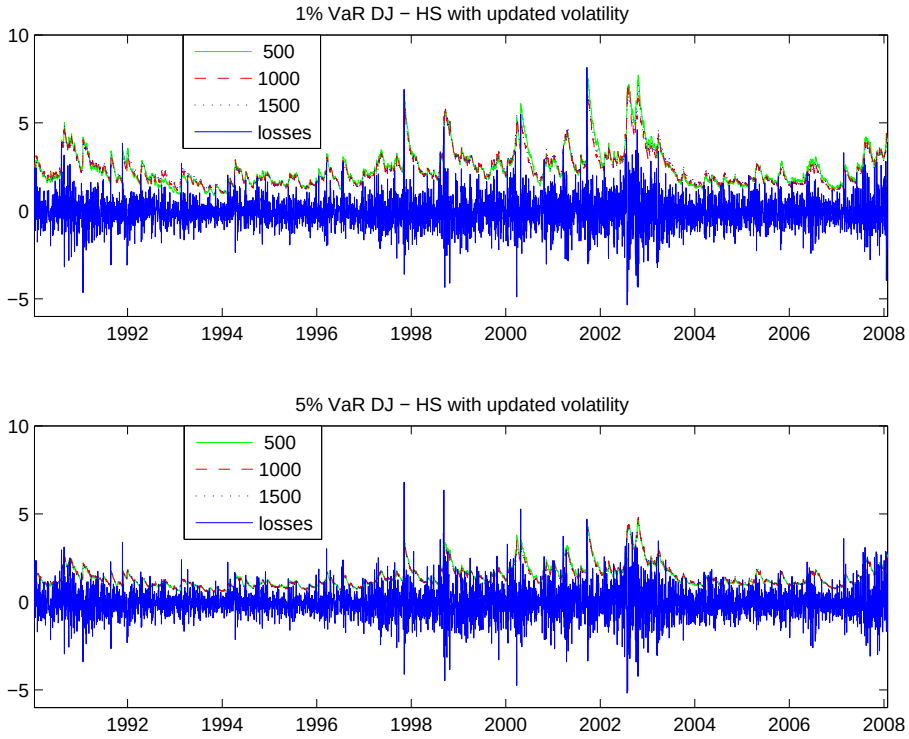


Figure A.17: The HS with updated volatility – VaR for Dow Jones and losses.

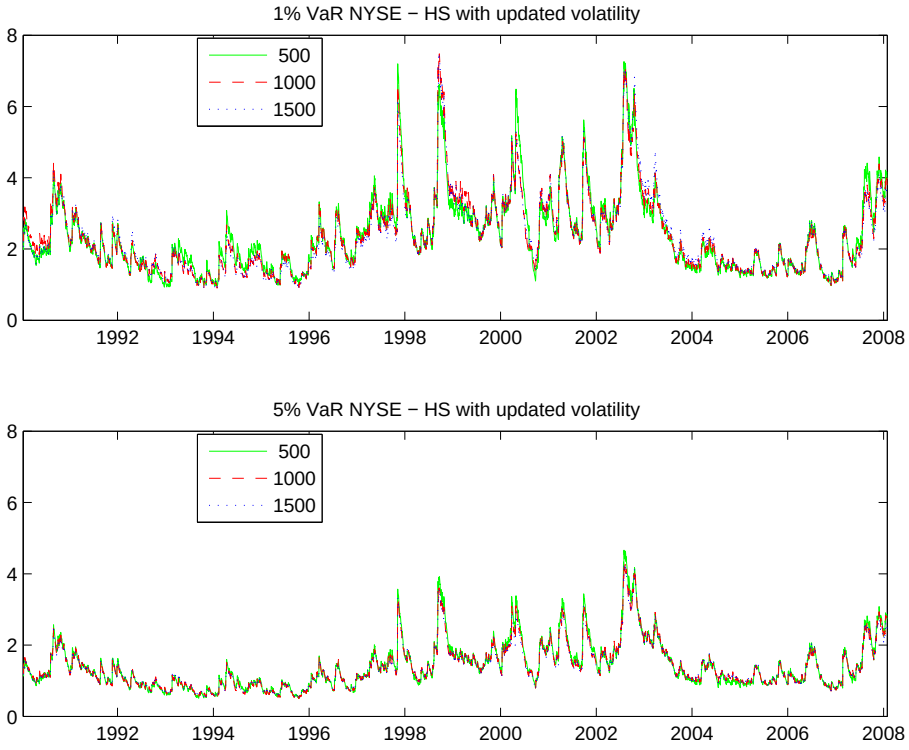


Figure A.18: The HS with updated volatility – VaR for NYSE.

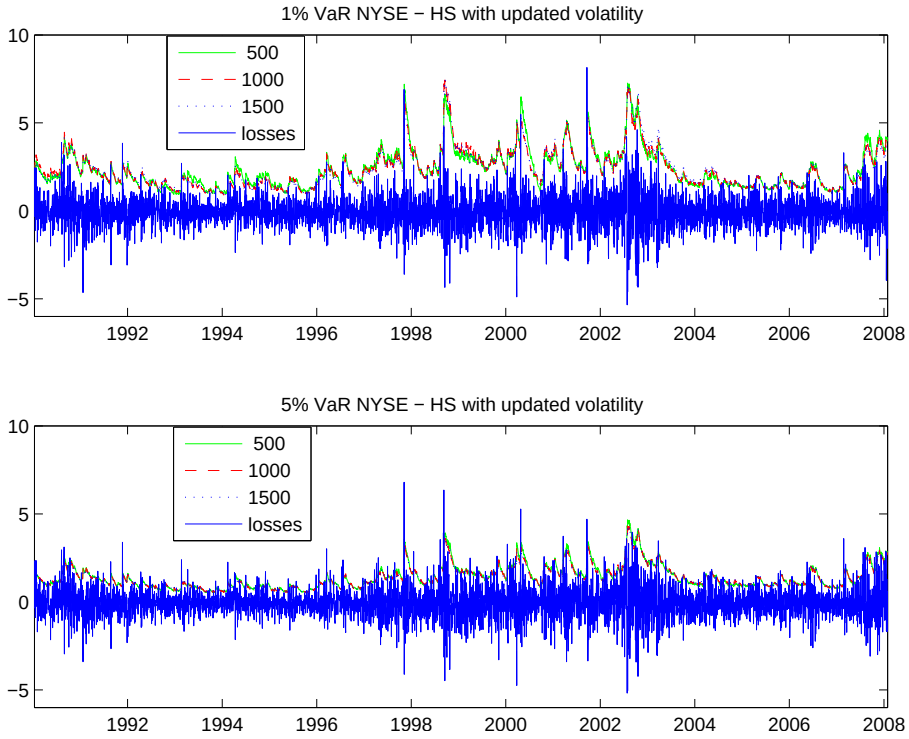


Figure A.19: The HS with updated volatility – VaR for NYSE and losses.

CAViaR models – 1% VaR – the first part.								
	Symmetric absolute value				Asymmetric slope			
	S&P	NASDAQ	DJ	NYSE	S&P	NASDAQ	DJ	NYSE
Beta1	0,104	0,338	0,146	0,172	0,169	0,162	0,313	0,199
Standard errors	0,038	0,031	0,012	0,018	0,032	0,037	0,063	0,081
p-values	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,007
Beta2	0,932	0,811	0,909	0,890	0,861	0,809	0,754	0,806
Standard errors	0,021	0,010	0,010	0,023	0,020	0,042	0,051	0,056
p-values	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Beta3	0,177	0,467	0,211	0,272	*-0,031*	0,295	*0,036*	*0,060*
Standard errors	0,016	0,018	0,029	0,093	0,072	0,145	0,062	0,176
p-values	0,000	0,000	0,000	0,002	0,337	0,021	0,282	0,366
Beta4					0,513	0,569	0,668	0,641
Standard errors					0,092	0,156	0,161	0,134
p-values					0,000	0,000	0,000	0,000
RQ	203,263	218,019	193,074	185,702	195,621	210,527	186,049	179,556
Hits in-sample(%)	1,026	1,026	1,008	0,990	1,008	1,008	1,026	0,990
Hits out-of-sample(%)	1,000	1,000	1,400	1,400	1,000	1,400	1,600	2,200
DQ out-of-sample	0,004*	0,928	0,042	0,962	0,004*	0,342	0,027	0,075

Table A.1: CAViaR models – 1% VaR – the first part. The first 5 554 observations were used to estimate the model and the remaining 500 observations were used for out-of-sample testing. Coefficients which are not significant at 5 % are *emphasized*; "*" denotes rejection according to the DQ test at 1% significance level.

CAViaR models – 1% VaR – the second part.								
	Indirect GARCH				Adaptive			
	S&P	NASDAQ	DJ	NYSE	S&P	NASDAQ	DJ	NYSE
Beta1	0,093	0,356	0,269	0,221	0,632	0,991	0,634	0,577
Standard errors	0,038	0,100	0,052	0,053	0,140	0,062	0,107	0,270
p-values	0,007	0,000	0,000	0,000	0,000	0,000	0,000	0,016
Beta2	0,932	0,799	0,853	0,858				
Standard errors	0,007	0,009	0,009	0,007				
p-values	0,000	0,000	0,000	0,000				
Beta3	*0,310*	*1,024*	*0,627*	*0,664*				
Standard errors	0,277	0,829	0,634	1,063				
p-values	0,132	0,108	0,161	0,266				
RQ	200,949	214,536	192,580	184,666	211,903	249,771	206,052	197,399
Hits in-sample(%)	1,026	1,026	1,026	1,026	0,954	0,972	0,918	0,936
Hits out-of-sample(%)	1,400	1,000	1,600	1,400	1,600	1,200	1,400	1,200
DQ out-of-sample	0,042	0,846	0,057	0,030	0,035	0,812	0,950	0,012

Table A.2: CAViaR models – 1% VaR – the second part. The first 5 554 observations were used to estimate the model and the remaining 500 observations were used for out-of-sample testing. Coefficients which are not significant at 5 % are *emphasized*; "*" denotes rejection according to the DQ test at 1% significance level.

CAViaR models – 5% VaR – the first part.								
	Symmetric absolute value				Asymmetric slope			
	S&P	NASDAQ	DJ	NYSE	S&P	NASDAQ	DJ	NYSE
Beta1	0,033	0,066	0,062	0,050	0,027	0,039	0,058	0,064
Standard errors	0,004	0,008	0,015	0,006	0,009	0,007	0,012	0,015
p-values	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000
Beta2	0,959	0,928	0,934	0,940	0,936	0,886	0,903	0,879
Standard errors	0,006	0,013	0,020	0,010	0,013	0,016	0,024	0,020
p-values	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Beta3	0,084	0,148	0,101	0,109	*0,017*	0,106	*0,005*	*0,014*
Standard errors	0,009	0,025	0,030	0,020	0,019	0,041	0,019	0,027
p-values	0,000	0,000	0,000	0,000	0,173	0,005	0,397	0,300
Beta4					0,178	0,308	0,239	0,320
Standard errors					0,026	0,035	0,065	0,054
p-values					0,000	0,000	0,000	0,000
RQ	615,202	750,075	591,234	566,362	604,475	735,579	576,583	550,288
Hits in-sample(%)	5,041	5,041	5,041	5,059	5,041	5,005	5,041	5,023
Hits out-of-sample(%)	6,400	6,200	6,800	7,200	6,400	5,600	6,000	6,200
DQ out-of-sample	0,083	0,199	0,014	0,111	0,126	0,183	0,130	0,050

Table A.3: CAViaR models – 5% VaR – the first part. The first 5 554 observations were used to estimate the model and the remaining 500 observations were used for out-of-sample testing. Coefficients which are not significant at 5 % are *emphasized*; "*" denotes rejection according to the DQ test at 1% significance level.

CAViaR models – 5% VaR – the second part.								
	Indirect GARCH				Adaptive			
	S&P	NASDAQ	DJ	NYSE	S&P	NASDAQ	DJ	NYSE
Beta1	0,018	0,019	0,041	0,036	0,362	0,339	0,324	0,378
Standard errors	0,010	0,009	0,010	0,011	0,043	0,070	0,031	0,059
p-values	0,041	0,016	0,000	0,001	0,000	0,000	0,000	0,000
Beta2	0,942	0,916	0,922	0,913				
Standard errors	0,007	0,006	0,003	0,007				
p-values	0,000	0,000	0,000	0,000				
Beta3	*0,120*	*0,224*	*0,139*	*0,167*				
Standard errors	0,197	0,207	0,126	0,247				
p-values	0,272	0,140	0,136	0,249				
RQ	615,924	746,317	588,687	565,685	615,537	762,598	592,620	565,260
Hits in-sample(%)	5,095	5,005	5,023	5,077	4,699	4,717	4,555	4,591
Hits out-of-sample(%)	6,400	5,400	6,400	6,400	5,600	5,600	6,000	5,600
DQ out-of-sample	0,107	0,131	0,078	0,511	0,419	0,194	0,012	0,822

Table A.4: CAViaR models – 5% VaR – the second part. The first 5 554 observations were used to estimate the model and the remaining 500 observations were used for out-of-sample testing. Coefficients which are not significant at 5 % are *emphasized*; "*" denotes rejection according to the DQ test at 1% significance level.

Bibliography

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance*, 9:203–228.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). VaR Without Correlations for Portfolios of Derivative Securities. *Journal of Futures Markets*, 19:583–602.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (2002). Backtesting Derivative Portfolios with Filtered Historical Simulation (FHS). *European Financial Management*, 8(1):31–58.

Bollerslev, T. (1986). Generalised Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Boudoukh, J., Richardson, M., and Whitelaw, R. F. (1998). The Best of Both Worlds. *Risk*, 11:64–67.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4):841–862.

Danielsson, J. and de Vries, C. G. (2000). Value–at–Risk and Extreme Returns. *Annales d'Economie et de Statistique*, 60:239–270.

Duffie, D. and Pan, J. (1997). An Overview of Value at Risk. *Journal of Derivatives*, 4(3):7–49.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics*, 22(4):367–381.

Gaivoronski, A. A. and Pflug, G. (2005). Value-at-Risk in Portfolio Optimization: Properties and Computational Approach. *Journal of Risk*, 7(2):1–31.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2:211–250.

He, X. and Wang, G. (1997). Convergence of Depth Contours for Multivariate Datasets. *The Annals of Statistics*, 25(2):495–504.

Huber, P. J. (1967). The Behaviour of Maximum Likelihood Estimates Under Nonstandard Conditions. *Proceedings of the Fifth Berkeley Symposium*, 4:221–233.

Huisman, R., Koedijik, K. G., and Pownall, R. A. J. (1998). VaR-x: Fat Tails in Financial Risk Management. Papers 54, Southern California - School of Business Administration.

Hull, J. and White, A. (1998). Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk. *Journal of Risk*, 1:5–19.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1):33–50.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Morgan, J. P. (1995). *RiskMetricsTM – Technical Document*. 3rd edition, J. P. Morgan.

Pritsker, M. (2006). The Hidden Dangers of Historical Simulation. *Journal of Banking and Finance*, 30:561–582.

Rachev, S. and Mittnik, S. (2000). *Stable Paretian Models in Finance*. John Wiley & Sons, Chichester.

Weiss, A. (1991). Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation. *Econometric Theory*, 7:46–68.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press, New York.